VIDEO QUALITY PREDICTION FOR VIDEO OVER WIRELESS ACCESS NETWORKS (UMTS AND WLAN)

A. Khan

Ph.D. December 2011

To our children, Humzah, Laith and Mariam

VIDEO QUALITY PREDICTION FOR VIDEO OVER WIRELESS ACCESS NETWORKS (UMTS AND WLAN)

by

Asiya Khan

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing and Mathematics
Faculty of Science and Technology

Partially sponsored by the EU FP7 ADAMANTIUM Project (EU Contract 214751)

December 2011

Video Quality Prediction for Video over Wireless Access Networks (UMTS and WLAN)

Asiya Khan

Abstract

Transmission of video content over wireless access networks (in particular, Wireless Local Area Networks (WLAN) and Third Generation Universal Mobile Telecommunication System (3G UMTS)) is growing exponentially and gaining popularity, and is predicted to expose new revenue streams for mobile network operators. However, the success of these video applications over wireless access networks very much depend on meeting the user's Quality of Service (QoS) requirements. Thus, it is highly desirable to be able to predict and, if appropriate, to control video quality to meet user's QoS requirements. Video quality is affected by distortions caused by the encoder and the wireless access network. The impact of these distortions is content dependent, but this feature has not been widely used in existing video quality prediction models.

The main aim of the project is the development of novel and efficient models for video quality prediction in a non-intrusive way for low bitrate and resolution videos and to demonstrate their application in QoS-driven adaptation schemes for mobile video streaming applications. This led to five main contributions of the thesis as follows:

(1) A thorough understanding of the relationships between video quality, wireless access network (UMTS and WLAN) parameters (e.g. packet/block loss, mean burst length and link bandwidth), encoder parameters (e.g. sender bitrate, frame rate) and content

type is provided. An understanding of the relationships and interactions between them and their impact on video quality is important as it provides a basis for the development of non-intrusive video quality prediction models.

(2) A new content classification method was proposed based on statistical tools as content type was found to be the most important parameter.

(3) Efficient regression-based and artificial neural network-based learning models were developed for video quality prediction over WLAN and UMTS access networks. The models are light weight (can be implemented in real time monitoring), provide a measure for user perceived quality, without time consuming subjective tests. The models have potential applications in several other areas, including QoS control and optimization in network planning and content provisioning for network/service providers.

(4) The applications of the proposed regression-based models were investigated in (i) optimization of content provisioning and network resource utilization and (ii) A new fuzzy sender bitrate adaptation scheme was presented at the sender side over WLAN and UMTS access networks.

(5) Finally, Internet-based subjective tests that captured distortions caused by the encoder and the wireless access network for different types of contents were designed. The database of subjective results has been made available to research community as there is a lack of subjective video quality assessment databases.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This study was funded in part by the ADAMANTIUM EU FP7 Project (EU Contract 214751).

Publications:

The original work presented in this thesis has been published in refereed Journals and International Conferences.

REFEREED JOURNAL

1. A. Khan, L. Sun and E. Ifeachor, "QoE Prediction Model and its Application in Video Quality Adaptation over UMTS networks", *IEEE Transactions on Multimedia (Accepted for publication).*
2. A. Khan, L. Sun and E. Ifeachor, "Learning Models for Video Quality Prediction over WLAN and UMTS Networks", *IET Communications*, Volume 4, Issue 12, pp 1380-1403, August 2010.
3. A. Khan, L. Sun, E. Jammeh and E. Ifeachor, "QoE-driven Adaptation Scheme for Video Applications over Wireless Networks", *IET Communications* Special Issue on *"Video communications over wireless networks",* Vol. 4, No.2, July, 2010.
4. A. Khan, L. Sun, J. Fajardo, F. Liberal, H. Koumaras and E. Ifeachor, "Video Quality Prediction Models based on Video Content Dynamics for H.264 Video over UMTS Networks", *International Journal of Digital Multimedia Broadcasting* special issue *"IP and Broadcasting Systems Convergence (IPBSC)"*, Vol. 2010, Article ID 608138, **http://www.hindawi.com/journals/ijdmb/2010/si.ipbsc.html**

5. E. Jammeh, I. Mkwawa, A. Khan, M. Goudarzi, L. Sun and E. Ifeachor, "Quality of Experience (QoE) Driven Adaptation Scheme for Voice/Video over IP", *Telecommunications Journal*, Springer Editions, Special Issue on the "*Quality of Experience issues in Multimedia Provision",* ISSN: 1018-4864 (Print) 1572-9451 (Online) DOI: 10.1007/s11235-010-9356-5, 9th June 2010.
6. A. Khan. L. Sun and E. Ifeachor, "Video Quality Assessment as Impacted by Video Content over Wireless Networks", *Journal on Advances in Networks and Services* **Journal Link** , Vol.2, No.2&3, Nov. 2009 - *Invited Paper*
7. A. Khan, L. Sun and E. Ifeachor, "Content-based Video Quality Prediction for MPEG4 Video Streaming over Wireless Networks", *Journal of Multimedia*, Issue 4, Vol.4, August 2009 (ISSN : 1796-2048) pp 228-239.

REFEREED INTERNATIONAL CONFERENCES

8. A. Khan, L. Sun, J. Fajardo, I. Taboada, F. Liberaland E. Ifeachor, "Impact of End Devices on Subjective Video Quality Assessment for QCIF Video Sequences", 3rd Int. Workshop on Quality of Multimedia Experience (QoMEX), 7-9 Sept., Mechelen, Belgium, 2011 (Acceptance rate 45%).
9. A. Khan, I. Mkwawa, L. Sun and E. Ifeachor, "QoE-driven Sender Bitrate Adaptation for Video Applications over IP Multimedia Subsystem", *IEEE ICC*, 5-9 June, Kyoto, Japan, 2011. (Acceptance rate: 38%)
10. A. Khan, L. Sun, J. Fajardo, F. Liberal and E. Ifeachor, "Impact of RLC losses on Quality Prediction for H.264 Video over UMTS Networks", *IEEE ICME*, 2010, Singapore, 19-23 July 2010. (Acceptance rate: 30%).
11. A. Khan, L. Sun, J. Fajardo, F. Liberal and E. Ifeachor, "An ANFIS-based Hybrid Quality Prediction Model for H.264 Video over UMTS Networks", *IEEE CQR*, 2010, Vancouver, Canada, 8-10 June 2010. (Acceptance rate: 47%)

12. A. Khan, L. Sun, J. Fajardo, F. Liberal and E. Ifeachor, "Video Quality Prediction Model for H.264 Video over UMTS Networks and their Application in Mobile Video Streaming", *IEEE ICC*, 2010, Cape Town, South Africa, 23-27 May 2010 (Acceptance rate: 38%).
13. A. Khan, L. Sun, E. Jammeh and E. Ifeachor, "Content classification-based and QoE-driven Video Send Bitrate Adaptation Scheme", *Mobimedia*, London, UK, 7-9 Sept 2009.
14. A. Khan, L. Sun and E. Ifeachor, "Content Classification Based on Objective Video Quality Evaluation for MPEG4 Video Streaming over Wireless Networks", *World Congress on Engineering, International Conference on Wireless Networks (ICWN),* 1-3 July 2009, London, UK. (Acceptance rate: 56.73%).
15. A. Khan, L. Sun and E. Ifeachor, "Content Clustering-based Video Quality Prediction Model for MPEG4 Video Streaming over Wireless Networks", *IEEE International Conference on Communications (ICC)*, 14-18 June, 2009, Dresden, Germany. (Acceptance rate: 33%).
16. A. Khan, L. Sun and E. Ifeachor, "Impact of Video Content on Video Quality for Video over Wireless Networks", *Fifth International Conference on Autonomic and Autonomous Systems (ICAS),* April 20-25, 2009, Valencia, Spain. Awarded with **BEST PAPER AWARD** (Acceptance rate varies from 28% to 42%).
17. A. Khan, L. Sun and E. Ifeachor, "An ANFIS-based Hybrid Video Quality Prediction Model for Video Streaming over Wireless Networks", *Second International Conference on Next Generation Mobile Applications Services and Technologies (NGMAST 2008)*, September 16-19, 2008, Cardiff, Wales, UK.
18. A. Khan, Z. Li, L. Sun and E. Ifeachor, "Audiovisual Quality Assessment for 3G Networks in Support of E-Healthcare", *Proceedings of the 3rd International*

Conference on Computational Intelligence in Medicine and Healthcare (CIMED 2007), July 25-27, Plymouth, UK, 2007.

Signed……………………………………

Date…………………………………….

Word Count (52,495)

Acknowledgments

I was once told that PhD is a long journey of transformation from a 'novice' researcher to a 'professional' researcher. The help and support of many made this journey possible for me.

First, I would like to thank my supervisor Dr Lingfen Sun for the freedom she gave me to pursue my research interests, for her patience, professional guidance and encouragement throughout this project and for the tremendous amount of time and effort she spent to ensure the high quality of my papers and this thesis. I would also like to thank my Director of Studies Professor Emmanuel Ifeachor for his guidance, critical comments, time and effort spent with discussing my work with me and giving me the best advice ever – publish! It was the advice given to me by him on my first day and continuous reminders ever since that motivated me to write the numerous papers as part of my PhD and gave his invaluable comments ensuring their high quality and this thesis, and for that I am forever grateful.

I would also like to thank the members of Signal Processing Multimedia Communications (SPMC) Research group: Dr Brahim Hamadicharef, Dr Zhuoqun Li, Dr Emmanuel Jammeh, Dr Is-Haka Mkwawa and Mr Mohammad Goudarzi for all the help, support and constructive discussions for this project. It has been my great pleasure to have been working in the SPMC group during my four and a half years of studying towards Ph.D. Through my Research Assistant position sponsored by the EU FP7 project Adamantium I have thoroughly enjoyed collaborating with Mr Jose-Oscar Fajardo, Dr Fidel Liberal and Dr Harilaos Koumaras. I have hugely benefitted from the collaborations and discussions with them. I would also like to take the opportunity to thank the numerous reviewers who reviewed my papers. Without their constructive comments I would not have been able to raise the standard of my publications.

I would also like to thank my parents, brother and sisters for their love and support and to my father for being a source of inspiration to me. I would also like to thank my husband for his encouragement, support and faith in me to undergo this huge task.

Lastly, this thesis is dedicated to our beautiful three children for their endless love, support and for bearing with me!

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Glossary

3G	Third Generation
3GPP	3G Partnership Project (UMTS)
3GPP2	3G Partnership Project 2 (UMTS)
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ACR	Absolute Category Rating
AM	Acknowledged Mode
AQoS	Application Quality of Service
AVC	Advanced Video Coding
BLER	Block Error Rate
CBR	Constant Bit Rate
CT	Content Type
DCR	Degradation Category Rating
DMOS	Degradation Mean Opinion Score
DPCH	Downlink Dedicated Physical Channel
ESTI	European Telecommunications Standards Institute
EURANE	Enhanced UMTS Radio Access Network Extension
FR	Frame Rate
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GW	Gentle Walking
IP	Internet Protocol
IPTV	Internet Protocol Television

ISO/IEC	International Organisation for Standardization/International Electronics Community
ITU	International Telecommunication Union
LBW	Link Bandwidth
LV	Linguistic Variable
MAC	Medium Access Control
MOS	Mean Opinion Score
MBL	Mean Burst Length
MPEG	Moving Pictures Experts Group
MSE	Mean Squared Error
NS-2	Network Simulator Version 2
NQoS	Network QoS
PCA	Principal Component Analysis
PDCP	Packet Data Convergence Protocol
PER	Packet Error Rate
PHY	Physical Layer
PDP	Packet Data Protocol
PDU	Packet Data Unit
PQoS	Perceptual Quality of Service
PSNR	Peak Signal to Noise Ratio
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QoS	Quality of Service
RLC	Radio Link Control
RM	Rapid Movement
RMSE	Root Mean Squared Error
RNC	Radio Network Controller

RTCP	Real Time Transport Control Protocol (IETF)
RTP	Real Time Transport Protocol (IETF)
RTT	Round Trip Time
SAD	Sum of Absolute Values
SBR	Sender Bitrate
SDU	Service Data Unit
SI	Spatial Information
SGSN	Serving GPRS Support Node
SM	Small Movement
SSIM	Structural Similarity Index
ST	Spatio-Temporal
SVC	Scalable Video Coding
TB	Transport Block
TCP	Transport Control Protocol
TFRC	TCP Friendly Rate Control
TI	Temporal Information
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Radio Access Network
VBR	Variable Bit Rates
VoD	Video on Demand
VoIP	Voice over Internet Protocol
WLAN	Wireless Local Area Network

Chapter 1

Introduction

The exponential growth of video services and applications over wireless access networks motivated the set up of this project. The scope of this Chapter is to present the motivations behind this project which led to four fundamental research questions in this area. Further, to outline the aims and objectives of the project and present the main contributions of this thesis. The arrangement of this Chapter is as follows. Section 1.1 presents the motivations behind the project. The research questions are given in Section 1.2. Section 1.3 presents the Project aims and objectives. The major contributions are summarized in Section 1.4. A brief overview and the organisation of the thesis are given in Section 1.5.

1.1 Motivations

Transmission of video content over wireless access networks of Wireless Local Area Networks (WLAN) and Third Generation Universal Mobile Telecommunication System (3G UMTS) is growing exponentially and gaining popularity (applications such as video streaming, video call, IPTV), and is predicted to expose new revenue streams for mobile network operators. Digital videos are now available everywhere – from handheld devices to personal computers. However, due to the bandwidth constraints of wireless access networks (low bandwidths for UMTS and bottlenecks for WLAN) Quality of Service (QoS) still remains of concern. This is because low video quality leads to poor QoS which in turn leads to reduced usage of the application/services and hence reduced revenues. In order to meet user's QoS requirement, there is a need to predict, monitor and if necessary control quality.

Video quality can be evaluated either subjectively or based on objective parameters. Subjective quality is the users' perception of service quality (ITU-T P.910) [1]. Mean Opinion Score (MOS) is the most widely used metric. The most reliable method is subjective quality, but it is expensive and consumes a lot of time and hence, the need for an objective method that produces results comparable with those of subjective testing. Objective measurements can be performed in an intrusive or non-intrusive way. Intrusive measurements require access to the source. They compare the impaired videos to the original ones. Full reference and reduced reference video quality measurements are both intrusive [2]. Quality metrics such as Peak-Signal-to-Noise-Ratio (PSNR), SSIM [3], Q-value [4], VQM [5] and more recently ITU-T J.247 [6] are full reference metrics. Non-intrusive methods (reference-free), on the other hand do not require access to the source video. Non-intrusive methods are either signal or parameter based. Non-intrusive methods are preferred to intrusive analysis as they are more suitable for on-line quality prediction and control.

Non-intrusive models provide an effective and practical way to achieve this. However, research on video quality modeling is still limited. For example, existing non-intrusive models do not take into account the impact of several important parameters, such as, video content which has an impact on video quality achievable under same network conditions.

Existing video quality prediction algorithms that consider the effects of distortions caused by the encoder are presented in [7, 8]-[11], the ones that consider network impairments are presented in [12, 13]-[14] and the ones that consider video content features only are given in [15],[16]. In addition, they are restricted over IP networks and do not consider the losses that occur in the access network. However, with the growth of video services over wireless access networks it is important to take into account impairments that occur in the access network. Work on video quality assessment [17, 18]-[19] has shown repeatedly that video quality is affected by distortions caused both by the encoder and access network. In addition, these impairments are very much content dependent.

Standardization efforts in VQEG and ITU-T on non-intrusive video quality assessment and modeling are currently under study by the Study Group 9 (SG 9). Recently ITU-T SG 9 has produced a draft version of the test plan for the hybrid perceptual/bitstream models [20] for IPTV and mobile video streaming applications. The objective of the hybrid project is to evaluate models that estimate perceived video quality from IP headers, bitstreams and the decoded video signal of short video sequences. Hence, there is a need to develop new models that consider impairments due to the encoder and access network for all content types either based on efficient statistical or neural network to predict/monitor video quality non-intrusively over wireless access networks of WLAN and UMTS for existing applications such as video streaming, IPTV, etc. to sustain and for emerging applications such as 3D TV, etc. to expand. The model should be efficient and light weight and suitable for all types of video content so that it can be implemented at the receiver side to monitor and predict quality and if appropriate control the end-to-end perceived quality. This thesis specifically addresses low resolution and sender bitrate videos over bandwidth restricted wireless access networks such as UMTS. UMTS is still used with low bitrates, however, the author acknowledges that in theory Long Term Evolution (LTE) is now under research and available though not on commercial scale. However, it is a matter of time that LTE will be a reality. In that situation, the work done in this thesis forms the basis of non-intrusive video quality assessment and prediction and the models developed here can easily be applied to bigger bandwidths and hence, better network conditions. They will need re-training only.

These models have applications in:

- prediction and monitoring of end-to-end video quality in an objective and non-intrusive manner on live wireless access networks and access network readiness for existing and new video services for all content types.

- optimisation of the quality of video services according to the changing network conditions and manage the utilisation of available resources – both content and network.
- for quality of service control (e.g. sender bitrate adaptation from the model).

1.2 Research Questions

This thesis seeks to address the following research questions:

Q 1) What are the relationships between perceived video quality, UMTS and WLAN access network impairments (e.g. packet/block loss, link bandwidth), relevant parameters associated with the video codec (e.g sender bitrate and frame rate) and video content type?

This led to a significant research to investigate the relationships between perceived video quality, wireless access network impairments (e.g. packet/block loss, mean burst length, link bandwidth), relevant parameters associated with video codec (e.g. sender bitrate and frame rate) and content type. A fundamental investigation of the impact of these parameters on perceived video quality is undertaken using both Mean Opinion Scores (MOS) obtained from subjective tests and MOS obtained from Peak-Signal-to-Noise Ratio (PSNR) conversion from [21] to obtain objective measure of video quality. This led to ranking of the QoS parameters in order of importance of its impact on video quality. Two codecs are used in this study – MPEG4 and H.264/AVC. H.264/AVC is the recommended codec by ITU-T for low bitrate video communication over access networks of UMTS and is also used in SG 9 draft test plan [20]. MPEG4 is the mandatory codec for WLAN access networks.

This work will be discussed in Chapter 3.

Q 2) How should the video contents be classified objectively?

There exist variety of video contents from fast moving (e.g. sports type) to slow moving news presentation (e.g. head and shoulder). This led to a fundamental research to investigate methods of classifying video contents objectively. In order to predict video quality for all content types, it is important to find a method to classify the contents. Two methods were looked at in detail. Firstly, the spatio-temporal (ST) features of raw video were extracted. Based on the ST features the contents were classified using a statistical tool. This method is used for the prediction of the non-intrusive models in this thesis. In the second method, MOS values resulted from impairments caused by the encoder and access network parameters were used in the same statistical function to classify video contents.

This work will be discussed in Chapter 4.

Q 3) How should the perceived video quality be measured/predicted non-intrusively and efficiently over wireless access networks of WLAN and UMTS?

Prior to finding a method that can predict video quality non-intrusively, the following two fundamental questions were addressed as:

> *What are the acceptable Packet Error Rates (PER) for all content types over WLAN access network and hence find the threshold at which the user's perception of video quality is acceptable?*

This question was addressed by finding the acceptable packet error rates for different types of content before quality degraded and hence a relationship between video quality in terms of the Mean Opinion Score (MOS) and packet error rate was established.

> *What is the minimum Sender Bitrate (SBR) requirement for all content types to meet communication quality for acceptable QoS (MOS > 3.5) in the absence of any access network losses?*

This question was addressed by finding the acceptable sender bitrate for different types of content before quality degraded and hence a relationship was found between video quality in terms of MOS and sender bitrates.

Once the relationship between video quality (in terms of MOS) and encoder related parameters and access network related parameters were established, this led to the following research question as:

> *How to predict video quality from a combination of parameters in the encoder and access network for different types of video contents non-intrusively ?*

This enabled the development of both regression-based and neural network-based models to predict video quality. Work presented in Chapter 5 describes the development of regression-based models over WLAN and UMTS access networks. Chapter 6 describes the neural network-based models developed for both access networks. The dataset used for model development was generated by both objective and subjective tests. The dataset generated by objective methods was based on simulation (NS2 and OPNET). Internet-based subjective tests were carried out based on existing VoIP tests.

Q 4) How should the perceived video quality metric be exploited to optimize and control end-to-end video quality?

This led to the proposal of a QoS-driven scheme for optimising content provisioning and network planning for video applications over WLAN and thus answering two fundamental research questions in this area as:

> *How content providers can provide the optimised video content matching according to user's QoS requirements?*
>
> *How network providers can best utilise existing network resources according to user's QoS requirements?*

This work is discussed in Chapter 7.

Chapter 7 also describes the application of model in QoS control over UMTS and WLAN access networks by proposing a new fuzzy sender bitrate adaptation scheme that is QoS-driven (driven by the proposed model).

1.3 Project Aim and Objectives

The main aims of the project are (1) to investigate and evaluate the impact of distortions caused by the encoder, access network parameters and content type that affect video quality over wireless access networks of WLAN and UMTS, (2) to classify the video contents objectively, (3) to develop novel and efficient reference-free models for the prediction of video quality over wireless access networks of WLAN and UMTS for all contents (classified using content classification method) and (4) to apply the developed models in quality optimization and adaptation mechanisms for perceptual QoS control for video over wireless access networks.

Specific objectives of the research are to:

- Undertake a fundamental investigation to quantify the impact of access network impairments (e.g. packet/block loss, mean burst length, link bandwidth), encoder impairments (e.g. encoder sender bitrate and frame rate) and video content type on perceived video quality in wireless access networks of WLAN and UMTS. Further, to identify parameters which will be used for video quality prediction.
- Classify video contents objectively. This led to the development of two video classification methods. Further, to choose a method for content classification to be used for video quality prediction.
- Develop novel and efficient non-intrusive video quality prediction models to predict perceived video quality from a combination of QoS parameter.

- Investigate the applications of the above models in areas such as video quality optimization and QoS control (e.g. sender bitrate adaptive QoS control) in wireless access networks of WLAN and UMTS.

1.4 Contribution of Thesis

The contributions of the thesis are as following:

1. A detailed understanding of the relationships between video quality, wireless access network impairments (e.g. packet/block loss, link bandwidth, mean burst length), encoder impairments (e.g. encoder sender bitrate and frame rate) and video content type was carried out. An understanding of the perceptual effects of the key parameters on video quality is important as it provides a basis for the development of efficient regression models and robust artificial neural network learning models. Two modern video codecs recommended by ITU-T for WLAN and UMTS are used as MPEG4 and H.264. The QoS parameters were ranked in order of importance.

(The associated publications are [22] - [26]).

2. Content type was found to be the most important QoS parameter. Therefore, a content classification method was proposed based on statistical tools. The classification is carried out by extracting spatio-temporal features of raw video. This method is then used in the non-intrusive video quality prediction models. A second method was also investigated which was based on classifying the video contents from MOS values obtained from impairments caused by parameters associated with the encoder and the access network.

(The associated publications are [27, 28]).

3. New models to predict video quality non-intrusively over wireless access networks of WLAN and UMTS are presented. The models are predicted from a combination of

parameters associated with the encoder, the access network and content type. The proposed models are developed from MOS values obtained objectively from PSNR to MOS conversion and subjective MOS values. Subjective database gave good prediction accuracy (~93%) as compared to the results of the objective database (~87%). Efficient regression-based and neural network based models were developed for wireless access networks of WLAN and UMTS. The detailed contributions are:

- New non-linear regression models for predicting video quality based on a combination of parameters from the encoder, access network and content type are derived.

  (The associated publications are [29] -[32]).

- New learning models, based on Adaptive Neuro Fuzzy Inference System (ANFIS) for video quality prediction are given over wireless access networks of WLAN and UMTS. ANFIS is chosen as opposed to ANN as it combines the advantages of neural networks and fuzzy systems.

(The associated publications are [33, 34]-[35])

4. Two applications for the new perceived video quality prediction models are investigated.

   - The newly developed efficient regression models are applied for optimisation of content and network provisioning.

     (The associated publications are [36], [37])

   - A new fuzzy adaptation scheme at pre-encoding stage is proposed for QoS-driven adaptation of sender bitrate.

     (The associated publications are [32],[38],[39],[31])

5. Internet-based subjective tests were designed based on existing VoIP tests which allowed rapid assessment of video quality by subjects as recommended by the ITU-T Rec.500 [40] using the single stimulus Absolute Category Rating (ACR) method with a five point quality scale. The subjective results database has been made publicly available at [41] for research community as currently there is a shortage of video quality assessment database available that combines distortions caused by the encoder and access network for different types of video content.

(The associated publication is [42])

1.5 Outline of Thesis

The outline of the thesis is shown in Fig. 1.1 and described as follows:

Chapter 2 gives a brief background information on the protocols and architecture of wireless access networks of WLAN and UMTS. A detailed description of video quality measurement methods are also described in Chapter 2. Section 2.2 describes WLAN and UMTS system protocol and architecture. Access network performance and their characteristics are outlined in Section 2.3. Section 2.4 highlights coding techniques and codecs used. Video quality and the factors that affect it is presented in Section 2.5. Section 2.6 summarizes state-of-the-art video quality measurement methods.

Chapter 3 discusses the impact of QoS parameters on video quality. Section 3.2 presents the related work on video quality assessment. Section 3.3 outlines the objective experimental set-up used in this thesis over WLAN. Section 3.4 describes the impact of QoS parameters over WLAN. Section 3.5 describes set-up over UMTS. Section 3.6 describes the impact of QoS parameters over UMTS. Section 3.7 presents an analysis of the QoS parameters and 3.8 ranks the QoS in order of importance.

Chapter 4 presents two methods for video content classification. Section 4.2 presents related work. Video content dynamics are discussed in Section 4.3. Sections 4.4 and 4.5 presents the two methods used for video content classification.

Chapter 5 presents the reference free regression-based models over WLAN and UMTS access networks. Section 5.2 presents related work. The test set-up is presented in section 5.3. Section 5.4 presents the subjective tests over UMTS. The methodology is introduced in Section 5.5. The procedure for developing the models is outlined in Section 5.6. The regression-based models over WLAN are presented in Section 5.7 and over UMTS in Section 5.8. Section 5.9 presents model comparison and validation over external databases.

Chapter 6 presents the reference free ANFIS-based models over WLAN and UMTS. Section 6.2 presents the related work. Section 6.3 presents the background on ANFIS. Section 6.4 presents experimental set-up over UMTS and WLAN. ANFIS architecture, training and validation is presented in Section 6.5. Section 6.6 and 6.7 presents the ANFIS-based models over WLAN and UMTS. The model comparison and validation with external database is presented in Section 6.8.

Chapter 7 presents the application of the proposed regression-based model in optimization and adaptation. Section 7.2 presents the related work. Section 7.3 describes the optimization of content provisioning and network planning. Section 7.4 presents the QoS-driven adaptation scheme of sender bitrate over WLAN and UMTS access networks. Results of the adaptation scheme are presented in Section 7.5.

Chapter 8 reviews the achievements of the project, concludes the thesis and suggests future work.

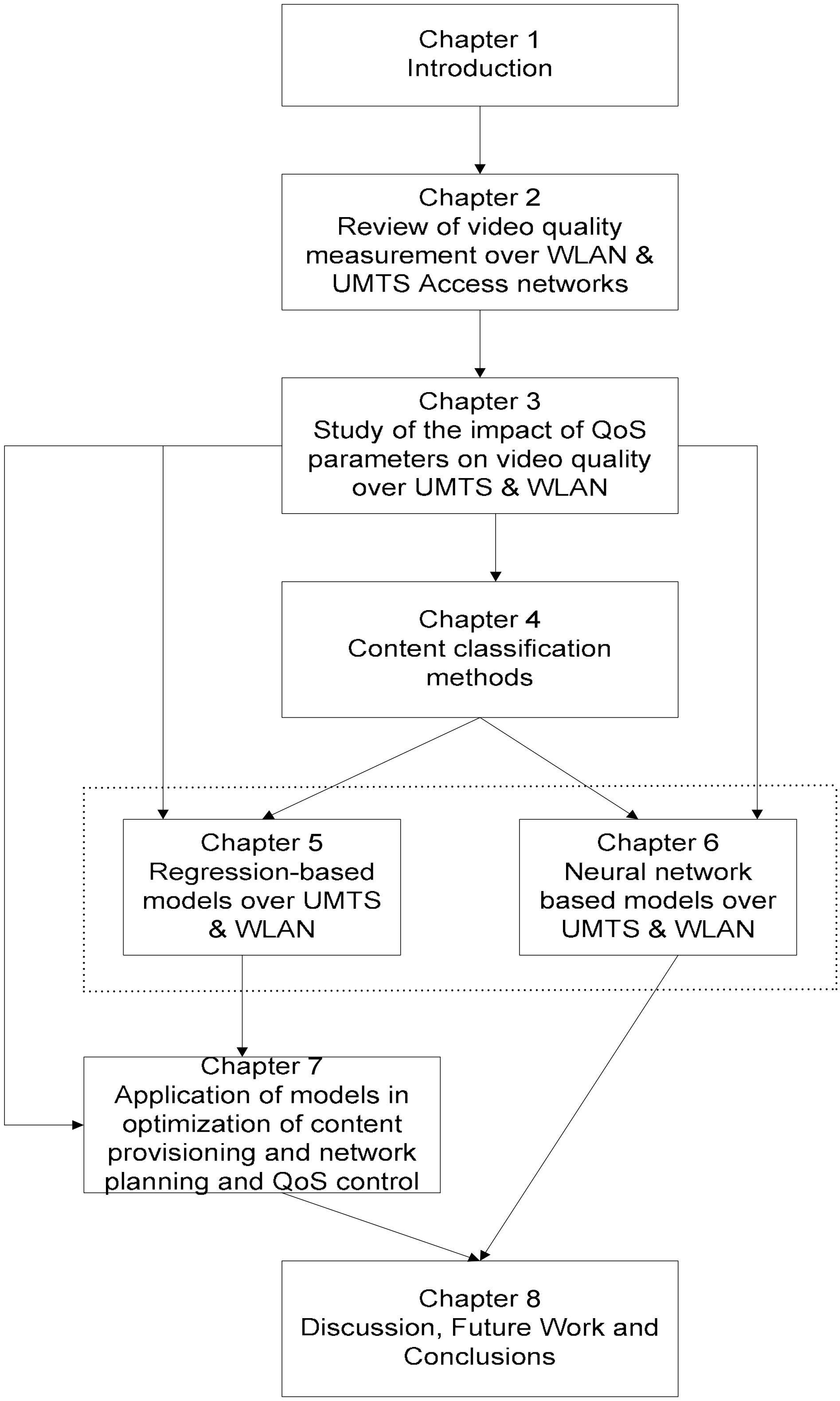


Figure 1.1. Outline of thesis

Chapter 2

Video quality measurement over WLAN and UMTS access networks

2.1 Introduction

With the advent of 3G mobile communication networks, the classical boundaries between telecommunications, multimedia and information technology sectors are fading. The goal of this convergence is the creation of a single platform that will allow ubiquitous access to Internet, multimedia services, interactive audiovisual services, and in addition (and most important) offering the required/appropriate perceived quality level at the end user's premises.

The aim of this chapter is to present a background on video content transmitted over WLAN and UMTS access networks and its Quality of Service issues, which underpins the work presented in later chapters. Section 2.2 summarizes the UMTS and WLAN system protocol and architecture. Section 2.3 outlines the characteristics of the access networks. A brief overview of video coding techniques and codecs used in this thesis is presented in section 2.4. The inter-relationships of QoS parameters over WLAN and UMTS access networks are presented in Section 2.5. A detailed review of work on video quality measurement techniques is presented in section 2.6, which serves as the foundations that the work of this thesis is based upon. Section 2.7 summarizes the chapter.

2.2 Video Transmission over WLAN and UMTS - System Architecture and Protocol

Figure 2.1 shows the conceptual transmission of video applications over WLAN (signalling part is not included) and UMTS. It consists of three parts - the sender, the access network (WLAN/UMTS) and IP core network and the receiver.

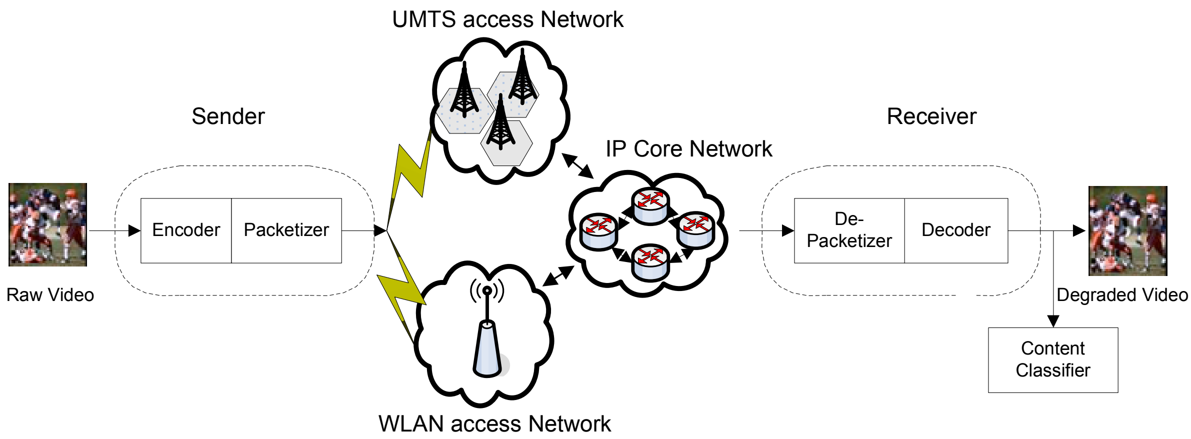


Figure 2.1. Video transmission over WLAN and UMTS access networks

At the sender, the raw video contents are first digitized and compressed by the encoder. After compression, they are packetized. This forms the payload part of a packet (e.g RTP packet). The packet is then formed by adding the headers (e.g. IP/UDP/RTP) to the payload which is sent to the wireless access networks. As the packet are sent over the wireless access networks, different network impairments (e.g. packet loss, delay, jitter, etc. due to access networks) may be introduced to the packets and further impairments due to IP networks. In this thesis we have considered the impact of access networks only. Video frames are extracted at the receiver by stripping off the packet headers from the payload by de-packetizer. There can be an added buffer which causes further delay. They are then decoded to recover the received video. Content classifier is at the receiver side. See details in Chapter 4.

In this thesis, the wireless access networks are simulated using widely available simulation tools. WLAN and UMTS access networks are simulated using NS2 [43]. Further UMTS is simulated using OPNET [44]. This is done to establish whether there is a preferred simulation platform for UMTS. Two codecs are used – MPEG4 [45] for WLAN and H.264 [46] for UMTS. MPEG4 is the mandatory codec whereas, H.264 is recommended for low bitrate communication which the aim of this study. H.264 was chosen for its better efficiency, more control at the encoder and it's an evolving codec.

The next two sections give an overview of WLAN and UMTS system architecture followed by a brief description of RTP protocol used for video streaming application in this thesis.

2.2.1 WLAN system architecture

The IEEE 802.11 [47] are a set of specifications for wireless standards and they specify an "over-the-air interface between a wireless client and a base station (access point) as well as among wireless clients". There are two basic operating modes - as infrastructure and adhoc mode. In the infrastructure mode, clients are allowed to roam between access points, however, roaming across routers is not allowed. The ad-hoc mode does not have an access point and individual nodes are allowed to participate in a peer-to-peer communication. In this thesis the infrastructure mode only is considered.

Three variations of the IEEE 802.11 standard are widely deployed. The 802.11a achieves 2.4GHz band has a maximum theoretical data rate of 11Mbps. It also operates on 1, 2 and 5 Mbps. The 802.11a and g will achieve the 5GHz and 2.4 GHz bands respectively. They both have a theoretical data rate of 54Mbps. However, if different modulation schemes are used then they can also achieve the lower data rates of 6, 10, 12, 18, 36 and 48Mbps. In this thesis, the simulations are based on IEEE802.11b with 11Mbps. IEEE801.11b was first wireless networking network that was widely adopted. Following, IEEE802.11b are the IEEE802.11g (54Mbps) and IEEE801.11n (600Mbps) standards. IEEE802.11n is still under development.

The mapping between Open Systems Interconnection (OSI) and IEEE 802 protocol layers is given in Fig. 2.2. OSI has seven layers as Physical, Data Link, Network, Transport, Session, Presentation and Application. The services and protocols specified in IEEE 802, map to the lower two layers of the seven-layer OSI networking interface model which are Data Link and Physical layers. In fact, IEEE 802 splits the OSI Data Link layer into two sub-layers named Logical Link Control (LLC) and Media Access Control (MAC).

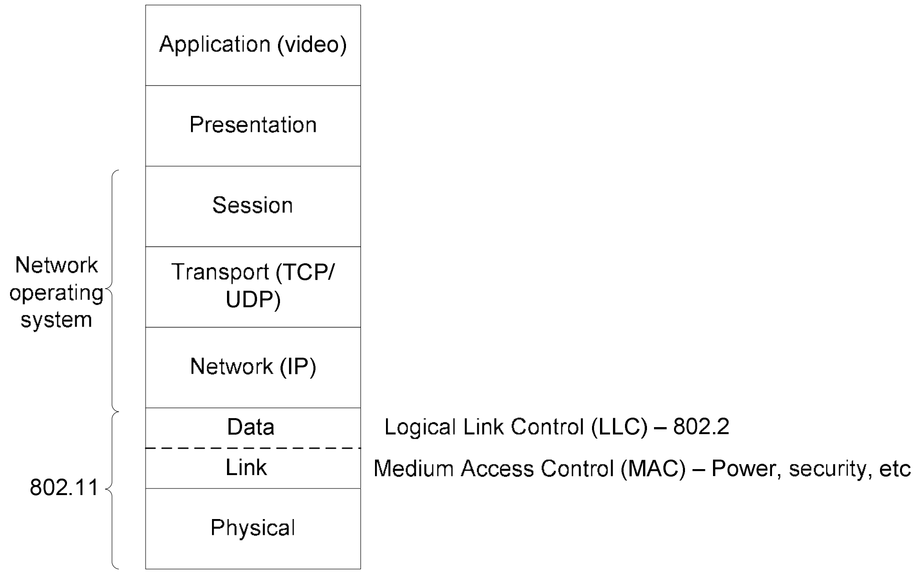


Figure 2.2. IEEE 802 protocol layers compared to OSI model

All of the components in the 802.11 architecture fall into either the media access control (MAC) sub layer of the data-link layer or the physical layer.

2.2.2 UMTS system architecture

3G (Third Generation) Systems [48] are aimed to provide a universal portability with different range of services that includes e.g., telephony, paging, messaging, Internet and broadband data. The ability for the transfer of information between access points is offered by UMTS teleservices and bearer services.. There are different QoS parameters for the bearer services for maximum transfer delay, delay variation and bit error rate. There are four QoS classes defined for UMTS network services: Conversational class (voice, video telephony, video gaming), Streaming class (multimedia, video on demand, webcast), Interactive class (web browsing, network gaming, database access), Background class (email, SMS, downloading).

The data rate targets are as follows:

Satellite and rural outdoor expects to achieve 144Kbps, 384Kbps is the target for urban outdoor, and 2048 Kbps is targeted for indoor and low range outdoor. All of these downlink data rates are the maximum theoretical values in each environment. The class of service dictates the actual data rates. E.g. they may vary from 32Kbps, for a single voice channel, to 768 Kbps in urban low speed connections.

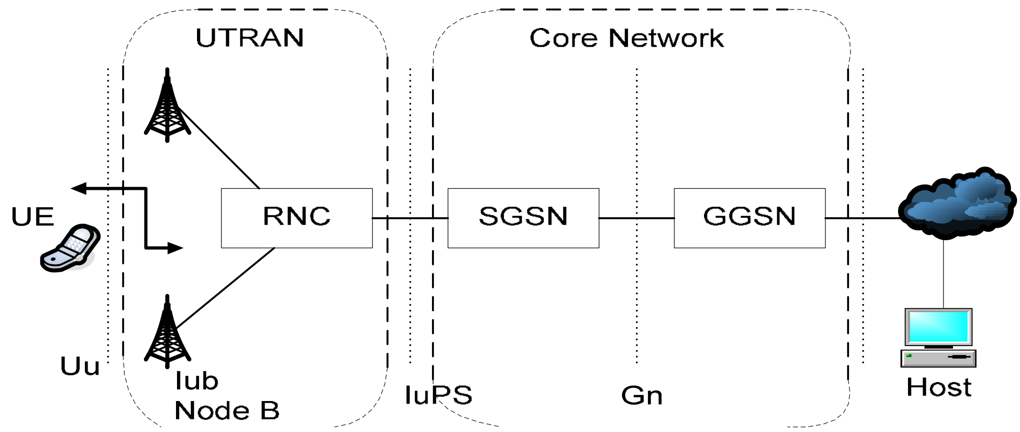


Figure 2.3. UMTS reference architecture

The UMTS architecture is shown in Fig. 3. It consists of one or many User Equipment (UEs), the UMTS Terrestrial Radio Access Network (UTRAN) and the core network. Node Bs of the UTRAN are connected to a Radio Network Controller (RNC). The base of the UMTS is the core network and it consists of the Serving GPRS Support Node (SGSN) and the Gateway GPRS Support Node (GGSN). The packets are routed to and from the UTRAN by the SGSNs, while GGSNs communicate with external IP networks. The UE is connected to Node B over the UMTS radio channel.

Fig. 2.3 shows the example of H.264 video being transmitted from a fixed Internet node to the mobile user. The packets come to the node B from the RNC and are then in a line or queue so that they can be divided down into smaller size packets. Every Packet Data

Convergence Protocol (PDCP) Packet Data Unit (PDU) is broken down into many Radio Link Controller (RLC) PDUs of fixed size. Each of these PDUs are then transmitted over the radio interface as a block. . Both the acknowledged (AM) and unacknowledged operation (UM) model is supported by the RLC model. The acknowledge modes (AM) assures delivery by retransmitting erroneous RLC blocks at the cost of transfer delay, whereas the UM provides not reliable but on time delivery of RLC blocks. For this reason, in this thesis only the AM mode is considered.

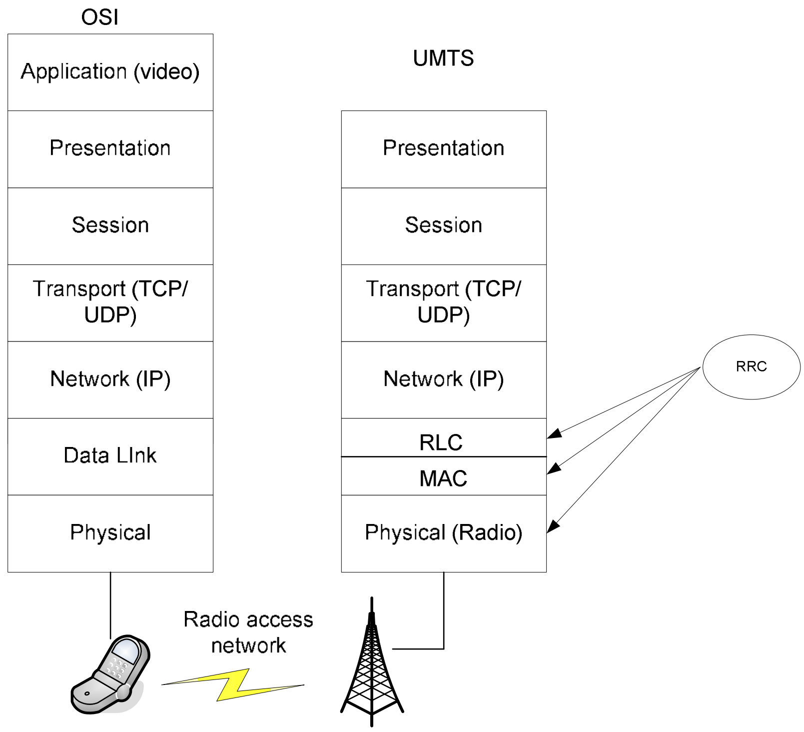


Figure 2.4. UMTS protocol layers compared to OSI model

Universal Mobile Telecommunications System (UMTS) is denoted as a 3rd generation cellular system and is developed within the framework that has been defined by the International Telecommunications Union (ITU) (known as the European version of IMT-2000). The process of standardising UMTS is carried out by the European Telecommunications Standards Institute (ETSI) in co-operation with other regional and national standardisation bodies around the globe to produce a standard that meets the needs of a growing market (UMTS Forum Report No. 2, 2000).

The UMTS protocol stack as compared to the OSI model is shown in Fig. 2.4. The physical layer is the radio signals, frequencies and channels. The data link layer is divided into Radio Link Control (RLC) and Medium Access Control (MAC). The network layer is the Internet Protocol (IP). The transport layer is the Transport Control Protocol (TCP) or the User Datagram Protocol (UDP). In UMTS there is also a Radio Resource Control (RRC) entity that manages the radio resources on several levels in the hierarchy. It has no corresponding entity in the OSI model.

2.2.3 Real Time Transport Protocol (RTP)

Real Time Transport protocol (RTP) [49] is used for most audio and video applications to transmit data . Existing transport protocols like UDP will run under RTP. The function of RTP is to provide applications that happen in real time with end-to-end delivery services such as payload type identification and delivery monitoring. The timing information of the data as received by the receiver from the sender can be reconstructed from the information provided by RTP. RTP messages contain a message sequence number which allows applications to detect packet loss, packet duplication, or packet reordering.

In an RTP message the RTP header is followed by the RTP payload. An RTP message of version 2 is shown in 2.5.

In Fig. 2.5, the content of the packet is added to the Payload Type (PT). The Service Description Protocol (SDP) describes the values for the codecs. The purpose of the Sequence Number is to find and reveal the lost packets and secondly to establish the order in which packets were sent to make loss detection easier. The sampling instant for the first octet of media data in a packet is in the timestamp. At the receiver, it can be used to help recover the clock frequency. If packets are lost then RTP does not include a loss recovery mechanism.

Payload Type (7 bits)	Sequence Number (16 bits)	Timestamp (32 bits)
Synchronisation Source identifiers	Contributing Source identifiers	Extension header

Figure 2.5. RTP Protocol stack

The extension of RTP is is given by the RTP control protocol (RTCP) [50] that will substitute member information in an on-going session. Data delivery is monitored by RTCP and the users are provided with some statistical functionality. RTCP can be used by the receiver as a feedback mechanism that informs the sender about the quality of an on-going session.

The transmission protocols for Mobile application are RTSP/RTP/UDP/IP. The Real Time Streaming Protocol (RTSP) is a network control protocol designed for use in entertainment and communications systems to control streaming media servers.

2.3 Access Network Performance Characteristics

2.3.1 Packet Loss and its characteristics over WLAN

When video is transmitted transmission over WLAN access network, the loss of packets is a major concern as it degrades the quality of video received at the receiver side. Packet are lost mainly if the access network is congested.

Markov processes are ideal to represent the packet loss behaviour of WLAN as there are many factors that cause packets to be lost and hence, packet loss is bursty in nature. There are many methods that can be found in literature that models this loss behaviour. In this thesis the 2-state Gilbert Model [51], [52] which will be discussed briefly in next section has been used with variable mean burst lengths. Parameters such as fading have not been explicitly considered, however, as part of NS2 [43] simulation environment, Rayleigh fading is implemented in Eurane [53] extension.

2-state Gilbert-Elliot Model

The 2-state GE model has been used recently in [4] over WLAN to represent packet loss characteristics. Fig. 2.6 illustrates a state diagram for a GE channel model. The GE model can be described as follows

"In the "good" state (G) errors occur with probability P_G while in the "bad" state (B) they happen with higher probability P_B. Also, P_{GB} is the probability of the state transiting from a good state to a bad state, and P_{BG} is the transition from a bad state to a good state".

The steady state probabilities of being in states G and B are given by Eq. (2.1):

$$\pi_G = \frac{P_{BG}}{P_{BG}+P_{GB}} \quad and \quad \pi_B = \frac{P_{GB}}{P_{BG}+P_{GB}} \tag{2.1}$$

respectively. The average packet error rate produced by the GE error model is given by Eq. (2.2).

$$Pavg = P_G \pi_G + P_B \pi_B \tag{2.2}$$

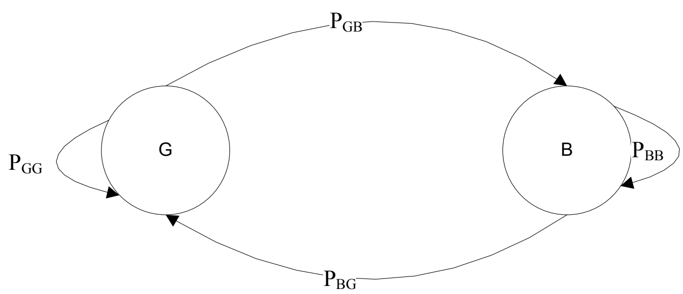


Figure 2.6. 2-state Markov model

2.3.2 Block Loss and its characteristics over UMTS

Block Error Rate (BLER) as opposed to packets lost is used as in UMTS networks. The transport blocks are passed from the physical layer to the Medium Access Control (MAC) layer together with an indication of error from Cyclic Redundancy Check. The output of the physical layer can be defined by the overall probability of the Block Error Rate (BLER). BLER is defined as the total number of blocks lost over total number of blocks sent.

Instead of setting up a target BLER value for the Packet Data Protocol (PDP) Context, the

UE model is modified in order to support the desired error characteristics. The implemented link loss model is special case of a 2-state Markov model [54] and its performance is provided by two parameters: the BLER and the Mean Burst Length (MBL). The 2-state Markov Model is depicted in Fig. 2.6. According to this model, "the network is either in good (G) state, where all packets are correctly delivered, or in bad (B) state, where all packets are lost as previously (WLAN)". Transitions from G to B and vice versa occur with probability $P_{GB} = 1\text{-}\beta$ and $P_{BG} = 1\text{-}\alpha$. Also, $P_{BB} = \alpha$ and $P_{GG} = \beta$. The average block error rate and mean burst length can be expressed as $MBL = (1\text{-}\alpha)^{-1}$ and $BLER = (1\text{-}\beta)/(2\text{-}\alpha\text{-}\beta)$. If $\alpha=0$, this reduces to random error model with the only difference that loss of two consecutive packets is not allowed.

2.3.3 Link Bandwidth (LBW)

Link Bandwidth (LBW) in UMTS is defined as the downlink bandwidth offered to the end customer. Currently, 384kbps is the maximum downlink bandwidth offered. Release 6 of 3GPP [48] offers LBW of up to 1-2 Mbps extending to 3-4 Mbps in the future. In this thesis LBW for UMTS is chosen as 128kbps, 256kbps and 384kbps. The downlink bandwidth is not increased beyond 384kbps to reflect the current 3G network. However, theoretical values especially with LTE are looking at 20Mbps and beyond. The work therefore, reflects worst case scenario of bottleneck in the access network. The LBW is simulated as the link between two routers in a WLAN set up. The LBW values are then used to distinguish this link. The LBW values are chosen from 32kbps-to 1Mbps over WLAN. In the WLAN set up higher link bandwidths are simulated (up to 1Mbps) as depending on the speed and location currently WLAN bottleneck access of up to 2Mps and beyond are possible.

2.4 Video Coding Techniques

2.4.1 Coding basic concept

Video compression algorithms ("codecs") manipulate video signals to dramatically reduce the storage and bandwidth required while maximizing perceived video quality. In general, video coding techniques are divided into two categories as lossless and lossy.

Lossless: This category will always make sure that the original data can be recovered exactly in its original format.

Lossy: This category states that the original data cannot be retrieved fully. The theory of coding is that small quantization errors in the high-frequency components of the image are not apparent by the human visual system.

In this thesis the focus is on lossy compression.

The progress of video coding research can be well represented via the development of video coding standards by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG) [55], [56].

Besides these coding standards there are also several significant progresses reported in the research literature. In this section an overview of the two codecs used in this thesis – MPEG4 for video transmitted over WLAN and H.264 for UMTS networks is provided.

2.4.2 Sampling and YUV Format

Both MPEG4 and H.264 use YUV format. YUV format separates a pixel into luma and chroma components. There is one luma (Y) component and two chroma components (U, V) per pixel. The luma component is associated with the brightness of the pixel while the other two chroma components are associated with the colour of the pixel. In this thesis YUV 4:2:0 format has been used. In YUV 4:2:0 format, every pixel has its distinct luma component while every 2x2 block of pixels share the same chroma (U, V) components. Since the human visual system is more sensitive to brightness (luma) than colour (chroma), so the chroma components are under sampled [55].

2.4.3 MPEG4

MPEG-4 video compression, as defined by the ISO/IEC Moving Picture Experts Group, is a standard for compressing and decompressing digital video aimed for distribution over media with limited bandwidth. The MPEG-4 standard suite includes 23 *parts* describing different aspects of multimedia compression and transfer. For instance, part 3 defines how to compress and encode audio to AAC-streams, and part 12, 14 and 15 describe the file and container formats. However, only two of the 23 parts describe how the raw video is encoded into a compressed data stream suitable for network transfer. These are part 2 and part 10. The latter describes the newest and most advanced codec, namely Advanced Video Coding (AVC), often referred to as H.264 (see later). The former is the one most commonly referred to as “MPEG-4 video”. This is a coding scheme with several different *profiles*, where the most normal are Simple Profile and Advanced Simple Profile (ASP). In the rest of this thesis, the term MPEG-4 will refer to the standardized MPEG-4 part 2 ASP.

With MPEG-compressing, none of the pictures in the raw-stream are stored or sent in its original form. Typically, the series of pictures or *frames* are divided into Groups Of Pictures (GOPs, see figure 2.7). The frames are then coded in various ways within each GOP. There are normally three types of frames in a MPEG GOP [57]:

Intra coded frames (I) are coded as single frames as in the case in JPEG. They do not have references to any other frames and are encoded independently of any other type of frames and are relatively large.

Predictive coded frames (P) are coded from an earlier I or P frame in the GOP They are typically smaller than the I-frames, but bigger than the B-frames.

Bi-directional coded frames (B) are coded on from preceding and succeeding I or P frames in the sequence. B-frames are typically smaller than both I and P-frames.

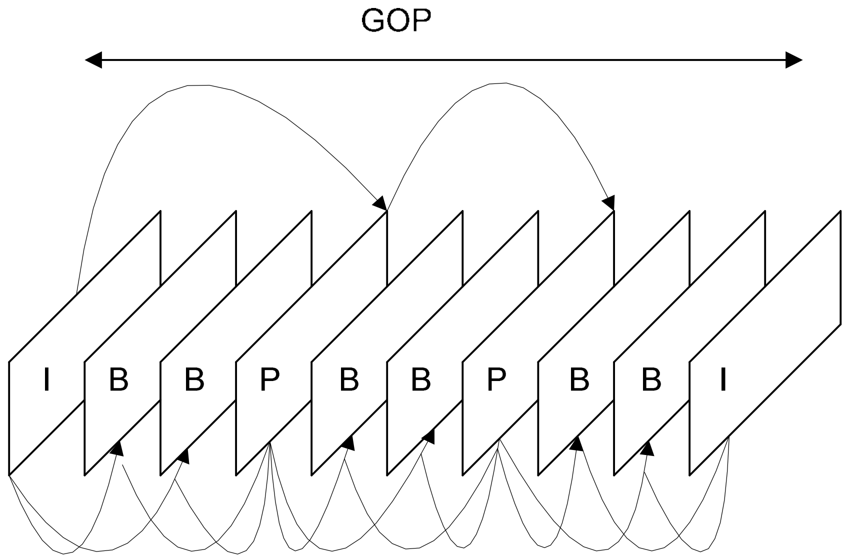


Figure 2.7. A sample of MPEG4 GOP (N=9, M=3)

A GOP pattern is identified by two parameters, GOP(N,M) – where N is the I-to-I frame distance and M is the I-to-P frame distance. For instance, as shown in Fig.2.7, G(9,3) would mean that one I frame, two P frames and six B frames are included in the GOP. The second I frame marks the beginning of the next GOP. Also the arrows in Fig. 2.7 show that the B frames and P frames that are decoded depend on the preceding or succeeding I or P frames.

The most important thing to note about this structure when talking about Quality of Service, is the fact that the I-frames are much more important to the quality of the video than the P-frames, which in turn are more important than the B-frames. Now, if an I-frame is lost during transmission due to congestion or erroneous links, the impact on the result is big. However, smart decoders are able to reconstruct a usable image even if one or several packets from one frame is lost as long as the packet containing the frame header is received. The number of packets needed to transfer one frame depends on the ratio between frame-size and the packet transfer unit size, called Maximum Transfer Unit (MTU). The overall result can be seen on measures such as the Peak Signal-to-Noise ratio (PSNR, see sub-section 2.6.4), Mean

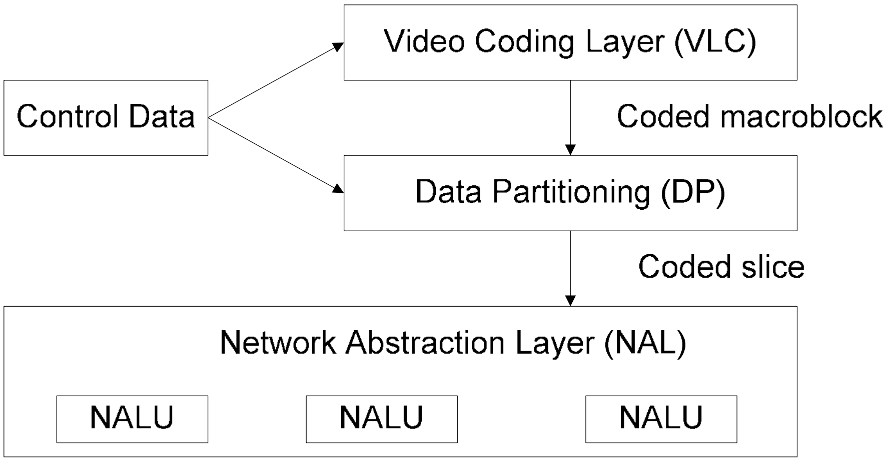
Opinion Score (MOS, see sub-section 2.6.1) or simply by watching the video in a MPEG-compatible player.

2.4.4 H.264/AVC Advanced Video Coding

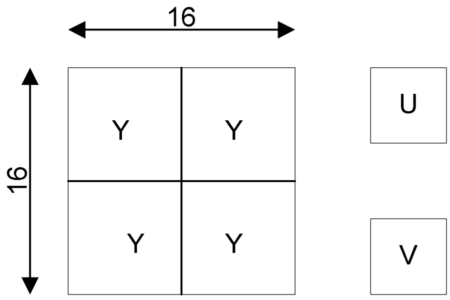
H.264/AVC is the latest codec and is also known as the next generation in the video compression technology in the MPEG4 standard. It is also known as MPEG4 part 10 and as Advanced Video Coding (AVC). The H.264 coded can produce the same quality as MPEG2 but at up to half the data rate. H.264 is the recommended codec for 3G to HD (High Definition) i.e. everything in between 40kbps to upwards of 10Mbps.

In this thesis H.264/AVC baseline profile (there are seven profiles recommended specific targeting classes of applications) has been used as it is designed for low complexity and low rate applications: (the target study of this thesis is mobile applications). The baseline profile does not support B slices. It only supports I and P slices. The design of H.264/AVC is depicted in Fig. 2.9a. From the figure, a Video Coding Layer (VCL) is designed to effectively represent the video content. The Network Abstraction Layer (NAL) configures the VCL illustration of the video and procures header information in away suitable for movement by a collection of transport layers or storage media. The NAL units (NALUs) are the output of NAL. The optional part is Data Partitioning (DP) This structure is given in Figure 2.8a.

A hybrid of temporal and spatial prediction is found in the VCL of H.264/AVC together with transform coding. Each frame is divided into non overlapping areas. These areas are called macroblocks (MB). They consist of 16×16 samples of the luma and 8×8 samples of each of the two chroma components. The macroblocks are arranged in slices which represent subsets of macroblocks that can be decoded independently. The macroblocks can be further divided into smaller blocks of up to 4 x 4 pixels. A 16 x 16 macroblock is shown in Figure 2.8b.



(a) H.264/AVC architecture



(b) A 16 x 16 macroblock showing YUV 4:2:0 sampling scheme [58]

Figure 2.8. Layer structure of H.264/AVC encoder

Slices

For compression, a video frame may be split into slices. A sequence of macroblocks that are grouped together is called a slice. A slice is coded and decoded independently of other slices in a frame. Typically, one encoded slice is packed in one NAL unit and transmitted in one packet. The advantage of using slices is that if a packet is lost then only a part of a video frame is lost instead of the whole frame. Since inter prediction between macroblocks of different slices is not allowed, use of slices decreases compression efficiency. H.264 standard defines the following slice types [58]:

I slices I or Intra slices contain macroblocks that are encoded using macroblocks in the same slice of the same frame. All the slices in the first frame of a video sequence are encoded as I slice.

P slices P or Predicted slices contain macroblocks that are encoded using macroblocks in a previously encoded and decoded frame. Some macroblocks in a P slice may be encoded in intra mode.

B slices B or Bi-Directional Predicted slices contain macroblocks that are encoded using macroblocks in the past and future I or P slices (in playback order). The decoding order of B slice is after the past and future I or P reference slices. In this thesis only IPPP structure is used.

2.5 Video Quality

Several factors influence video quality. These factors can be characterized as QoS parameters. In this thesis, the QoS parameters that are considered can be split into QoS affected by the access network and QoS affected by the encoder. The impact of the QoS parameters on video quality as perceived by the users is referred to as the Quality of Experience (QoE). QoE is difficult to measure as it goes beyond the boundary of measurable QoS parameters to how the users feel to a particular video service etc. For that reason, this thesis focuses only on the QoS parameters associated with the encoder (or application layer) and the access network (physical layer). The access network factors include access network packet/block loss, access network link bandwidth and other factors such as network jitter and network delay. These factors grouped together can then be defined as Network QoS (NQoS). The other factors of jitter and delay are not considered in this thesis. Delay is more significant in voice applications. Voice applications are real time and a delay in voice makes the user feel that the communication is lost. However, with video, delay is not as significant especially due to bigger cache memory now available. As the bandwidth's are increasing (LTE), delay will be even less significant. As the target application of this thesis is streaming video over wireless access networks, only packet/black losses were considered. The codec related parameters are sender bitrate and frame rate. The impact of both the encoder related and access network related parameters are very much dependent on the content type. Hence grouped together these factors can be defined as Application QoS (AQoS).

The characteristics of access network packet/block loss, MBL and LBW were discussed in detail in Section 2.3 earlier. The impact of the QoS parameters on perceived video quality will be discussed in Chapter 3.

2.6 Video Quality Measurement

Video quality can be evaluated either subjectively or based on objective methods as shown in Fig. 2.9. Subjective quality is the users' perception of service quality (ITU-T P.910) [1]. Mean Opinion Score (MOS) is the most widely used metric. The most reliable method of measuring video quality is the subjective way. However, this method consumes a lot of time and can be expensive. Hence, there a need for an objective method that produces results comparable with those of subjective testing. Objective measurements can be performed in an intrusive or non-intrusive way. Intrusive measurements require access to the source. They compare the impaired videos to the original ones. Full reference and reduced reference video quality measurements are both intrusive [2]. Quality metrics such as Peak-Signal-to-Noise-Ratio (PSNR), VQM [5], SSIM [3], Q-value [4] and J.247 [6] are full reference metrics. VQM and PEVQ are commercially used and are not publicly available. Non-intrusive methods (reference-free), on the other hand do not require access to the source video. Non-intrusive methods are either signal or parameter based. Non-intrusive methods are preferred to intrusive analysis as they are more suitable for on-line quality prediction/control.

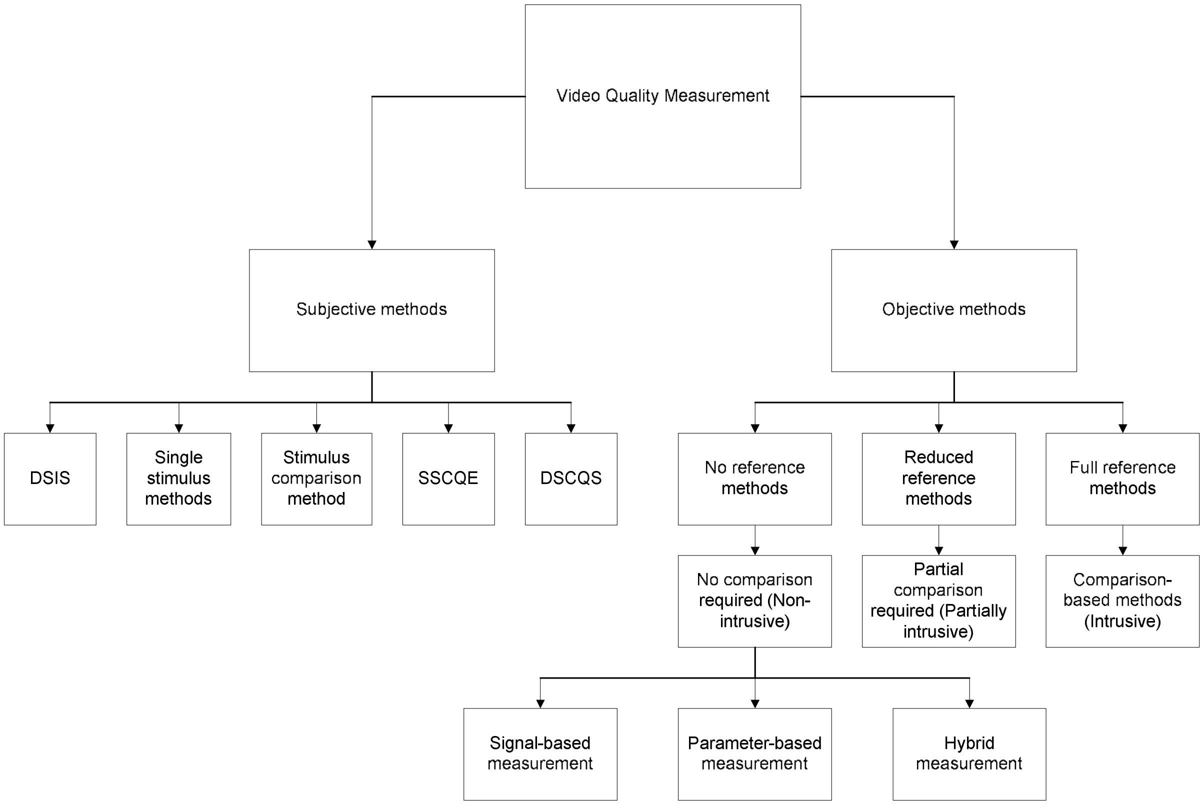


Figure 2.9. Classification of video quality assessment methods

2.6.1 Subjective Video Quality Measurement

International Telecommunication Union (ITU) and Video Quality Experts Group (VQEG) have defined the subjective methods such that a number of viewers (subjects) are selected who then watch the video clip in question under controlled conditions. After they have watched the clip, they then score the video clip. This score is called the Mean opinion Score (MOS) and has a value from 1 to 5.

Subjective test methods are described in ITU-R T.500-11 (2002) [40] and ITU-T Rec. P.910 (1999) [1]. These methods advise on the type of viewing conditions, the benchmark for observers and test material selection, the procedure to assess and statistical analysis methods. ITU-R Rec. BT.500-11 described subjective methods that are specialized for television applications, whereas ITU-T Rec. P.910 is intended for multimedia applications.

The most known and widely used subjective methods are:

- **Double Stimulus Impairment Scale (DSIS)**
- The subjects in this method are shown pairs of the degraded video clips along with many reference clips. The reference clip is shown before any degraded pair. Subjects score according to a scale of impairment as, "imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying". This scale is also known as the 5point scale. 5 is referred to as being imperceptible and 1 is very annoying.
- **Single Stimulus Methods** – The subjects in this method as opposed to the previous one are shown multiple separate scenes. This method can be run in one of two ways. : The first one is the single stimulus when the test scenes are not repeated and the second is the single stimulus with repetition of many times of the test scenes. As opposed to the previous one, three different types of methods for the scoring are used:
    - Adjectival: This is the same as described in Double Stimulus Impairment Scale with the difference that half scales are allowed.
    - Numerical: In this one, an 11-grade numerical scale is used. This is useful if a reference is not available.
    - Non-categorical: In this one, a continuous scale is used with no numbers. Alternately a large range of numbers can be used, e.g. 0 - 100.
- **Stimulus Comparison Method**: This method is used when two well matched monitors are available. The differences between pairs of scenes are scored in one of two ways:
    - Adjectival: This is a 7-grade scale labelled from +3 to -3 :The scale is translated as, "much better, better, slightly better, the same, slightly worse, worse, and much worse".

  - Non-categorical: This is very similar to the previous one where a continuous scale is used without numbers or a relation number used either in absolute terms or related to a standard pair.

- **Single Stimulus Continuous Quality Evaluation (SSCQE):** According to this method, the subjects watch a program of typically 20–30 minutes with no reference to the original video clip. The subjects rate using a slider continuously perceived quality at that instant in time on a scale from ‘bad’ to ‘excellent’. This corresponds to an equivalent numerical scale from 0 to 100.

  Double Stimulus Continuous Quality Scale (DSCQS): At DSCQS the viewers watch multiple pairs of quite short (i.e. 10 seconds) reference and test sequences. Each pair appears twice, with random order of the reference and the test sequence. The viewers do not know of the reference video clip order and are asked to rate each of the two separately on a continuous quality scale. This scale range from ‘bad’ to ‘excellent’. This is equivalent to a numerical scale from 0 to 100. The aforementioned methods are described in detail in the ITU-R Rec. T.500-11 document and are mainly intended for television signals. Based on slight modifications and adaptations of these methods, some other subjective evaluation methods like the Absolute Category Rating (ACR) and Degradation Category Rating (DCR) for multimedia services are described in ITU-T Rec. P.910 [1] and are listed below.

The Absolute Category Rating (ACR) method is the most commonly used method and it gives the Mean Opinion Score (MOS) as a metric of measurement. The Degradation Category Rating (DCR) method is also used in some situations and it gives the Degradation Mean Opinion Score (DMOS) as a metric.

2.6.1.1 Absolute Category Rating (ACR) method

In the ACR method, the viewers watch a video clip without watching the original reference clip. Once they have watched the clip, they are asked to give a quality rating. The quality rating is based on an opinion scale as shown in Table 2.1. The mean Opinion Score (MOS) is then calculated as an average of all opinion scores of the subjects.

Table 2.1 Opinion scale for ACR test

Category	Video Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Fig, 2.10 shows the time pattern for the presentation of video. The time of voting is equal to or less than 10ss.

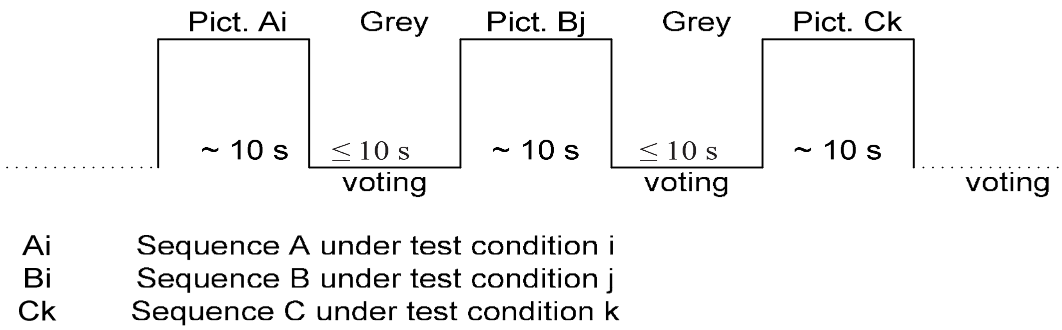


Figure 2.10. Stimulus presentation in the ACR method

2.6.1.2 Degradation Category Rating (DCR) Method

The DCR method is more effective when the quality difference in a video clip is minimum. The ACR method struggles to pick up the slight differences I quality (e.g. between 3 and 4). The DCR method is then recorded in an annoyance scale and a quality reference. The annoyance or degradation level is rated by the viewers by comparing the degraded video sequence to the original (reference). The rating scales or the degradation levels are shown in Table 2.2.

Table 2.2 Opinion scale for DCR test

Category	Video Quality
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Figure 2.11 shows the time pattern for the presentation of the video clips. The time of voting should be less than or equal to 10 seconds.

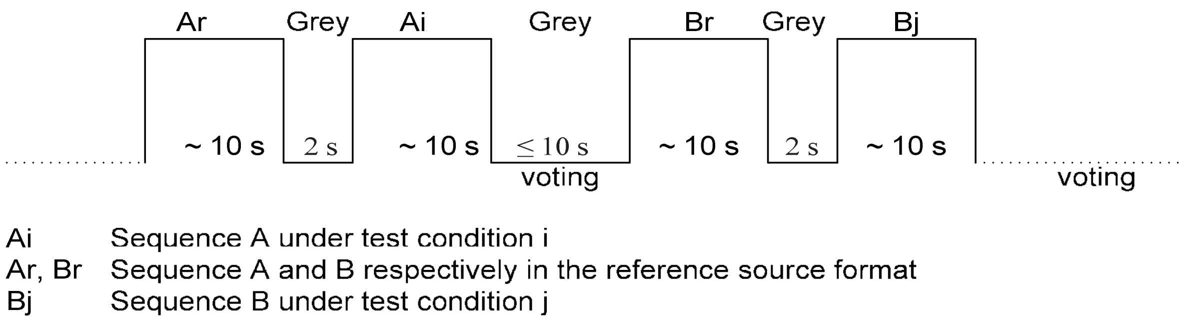


Figure 2.11. Stimulus presentation in the DCR method

2.6.1.3 Pair Comparison Method (PC)

In pair Comparison method, pairwise comparison of test sequences are repeatedly conducted. In comparison to other methods such as single stimulus and double stimulus, the test sequences are combined in all possible combinations, all pair of sequences are presented in both possible orders and subjects only need to provide preference between each pair instead of assigning a discrete or continuous score. The time pattern for the presentation of video clips can be shown by Figure 2.12. As previously, the voting time should be less than or equal to 10 s.

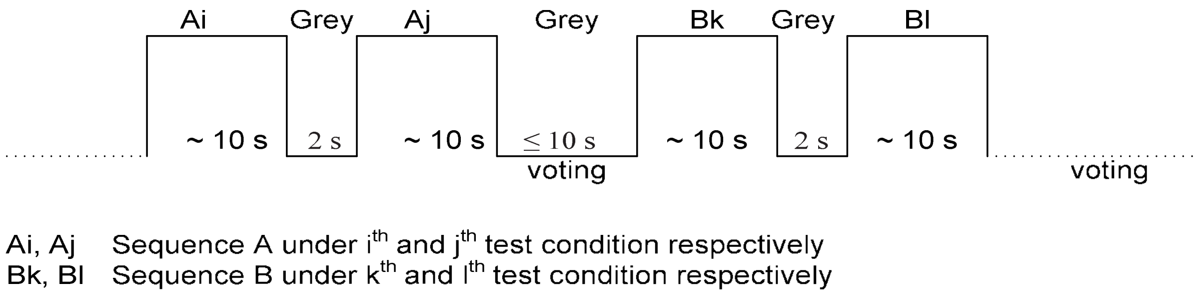


Figure 2.12. Stimulus presentation in the PC method

2.6.2 Experimental design of subjective tests

Once the data has been collected, several aspects must be taken into account in the experimental design of subjective tests. A description of these aspects is done in the next paragraphs. Subjective tests designed for this project were based on existing VoIP test in [59] and took into account all of the aspects described below. (See Chapter 5, section 5.4 for details).

Scene characteristics

It is important to select the test sequences representative of the data to be collected. Also, the choice of scenes should be different so the subjects are not bored. The test sequences should be the same for all subjects (scene characteristics is the same across the test) It. The short sequences should be less than 10 s and long sequences more than 10 s .

Replications

The replication of the video sequences is required by ITU-T P.910. At least two, otherwise three or four repetitions of the same test sequence should be shown to the viewers. This is important as it validates both subjects and the results produced by them.

Presentation order

The presentation order of the video sequences should be randomized. It should be different to the number of subjects in the same test. However, when analysing the results, presentation order should be taken into account. This is because if the viewers viewed a ‘bad’ (degraded) clip then viewed a ‘fair’ clip, they may rate is as ‘good’.

Viewers

The number of viewers recommended by ITU-T are from 4-40. In practice, a minimum of 15-18 non-expert viewers should be used for obvious reasons. Ideally they should be naïve viewers with little or no experience of these types of tests.

Viewing conditions

The viewing conditions should be uniform for all viewers, e.g. display equipment, seating position, etc.

Instructions to viewers

The subjects should be briefed about the intended application of the test to be undertaken. These instructions must be in writing to explain fully what is required from the viewers.

Training session

A training session should be included before any real testing is carried out. The purpose of this session is to familiarise the users to the type of test..

Evaluation

As seen before in the test methods description part, different kinds of evaluative scales are used depending on the test method used. Grading scales (five-graded, seven-graded for comparison or even with more points) or continuous scales are possible, and the scale must be clearly illustrated during the training phase or before. In addition, the viewers can be given a specially designed questionnaire to rate the test.

Apart from that, it is important to decide if the evaluation of the video is concentrated on the whole video or in concrete objects. In the case of short video tests, the assessment of quality can be applied to consider the impact of the quality of single objects on the overall quality of the video clip, contrary to long videos that should be evaluated in general perspective.

Subjective tests are composed of two phases: initial phase (instructions and training phase) and test sessions. A session should last less than half an hour. The sessions must be split up to provide for breaks so that the viewers are not tired or fatigued from the experience. This general structure is explained in Figure 2.13.

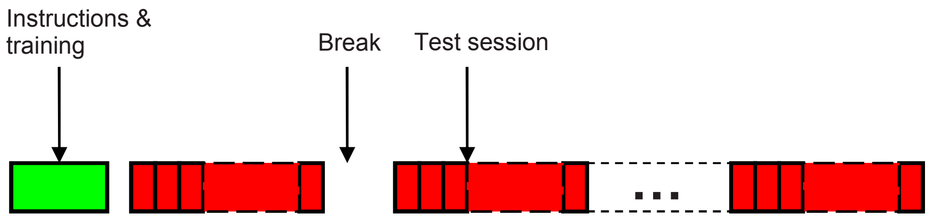


Figure 2.13. Subjective tests – general structure

2.6.3 Objective Quality Evaluation Methods

The subjective tests are expensive and take a lot of time due to the reasons explained above. This limits their implementation especially for research purposes.

This makes objective methods high desirable. VQEG SG9 [20] is dedicated to finding effective objective methods that can measure video quality comparable to that of subjective standards. Video quality measurements are divided into three main areas as full-reference, reduced-reference, and no-reference. Full-reference and reduced reference are intrusive measurements and are described in sub-sections 2.6.4.1 and 2.6.4.2. No-reference is a non-intrusive measurement and is described in Section 2.6.5.

2.6.4 Intrusive Video Quality Measurement Methods

Intrusive video quality measurement methods can be defined as full reference and reduced reference.

2.6.4.1 Full Reference Method

The full reference methods require reference to the source video. They compare the source video to the received degraded video and then provide an objective quality value. They are impractical for online monitoring and prediction and access to the source (original) video may not be possible.

The basic block diagram of the full reference method is depicted on Fig. 2.14

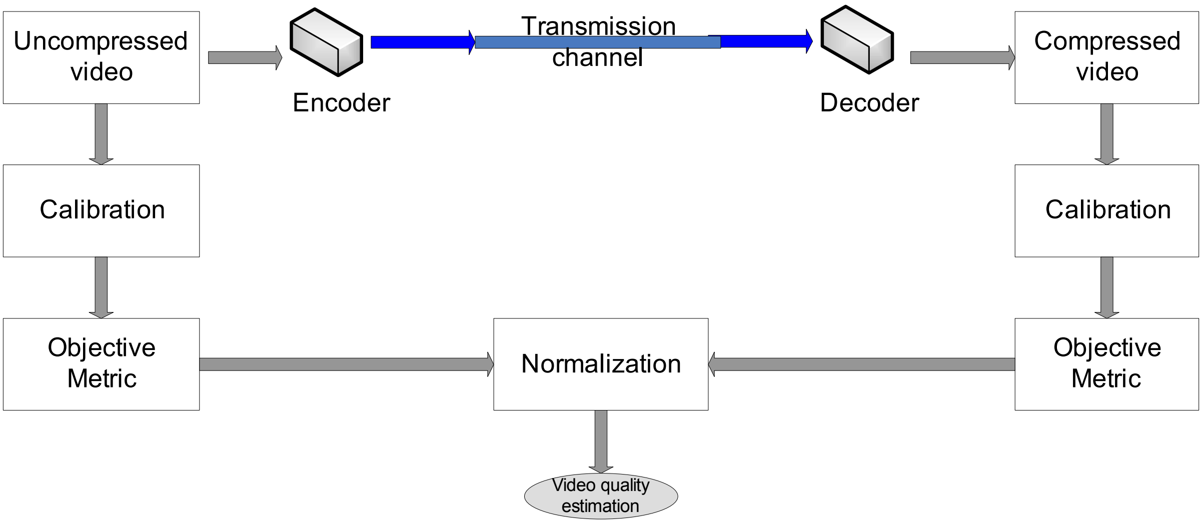


Figure 2.14. Full Reference Methods with single channel

PSNR, SSIM, VQM, Q-value and J.247 are all full reference methods. In this thesis PSNR has been used widely, followed by Q-value and SSIM for comparison.

PSNR

Peak-Signal-to-Noise-Ration (PSNR) is a well known objective video quality metric. It is most commonly used in video quality evaluation. It is a full reference metric where the PSNR value is measured by comparing the original frames to the reconstructed frames of the video sequence. PSNR is easily defined by the Mean Squared Error (MSE) of two images as shown by Eq. 2.3.

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \qquad (2.3)$$

The PSNR is defined in Eq. (2.4). The PSNR value approaches infinity when MSE is close to zero. Hence a higher PSNR value gives a higher video quality. Whereas, a small PSNR value high numerical differences between frames and hence low video quality.

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \end{aligned} \tag{2.4}$$

MAX_I is the maximum possible pixel value of the image in Eq. (2.4).

SSIM

The Structural Similarity Index Measurement (SSIM) provides a quality index measure of the similarity between two images. It was developed by Wang et al [3]. AS PSNR and MSE do not accurately depict human perception, SSIM was designed to represent the human perception and improve on methods like the PSNR.

The SSIM metric is calculated using Eq. 2.5. The measure between two images *x* and *y* of common size $N \times N$ is given in Eq. (2.5).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2.5}$$

This formula is applied only on luma to evaluate the image quality with a maximum value of 1 showing excellent quality.

Structural dissimilarity (DSSIM) is a distance metric derived from SSIM and is given in Eq. (2.6).

$$DSSIM(x,y) = \frac{1}{1 - SSIM(x,y)} \tag{2.6}$$

Q - value

The Q-value – defined as the decodable frame rate is defined as the number of frames that can be decoded at the receiver over the total number of frames originally sent by a video as given in Eq. (2.7).

$$Q = \frac{N_{dec}}{(N_{total-I} + N_{total-P} + N_{total-B})} \qquad (2.7)$$

Where Ndec is the summation of Ndec-I, Ndec-P and Ndec-B defined in [60]. A large Q value means that the video quality perceived by the end user is good.

2.6.4.2 Reduced Reference Methods

As opposed to full reference methods reduced reference methods use only some features extracted from the original video clip. The reduced reference method is shown by the block diagram in Fig. 2.15. Therefore, if a reduced-reference method is to be used in an IPTV system, the user requirement should specify the side channel, through which the feature data is transmitted.

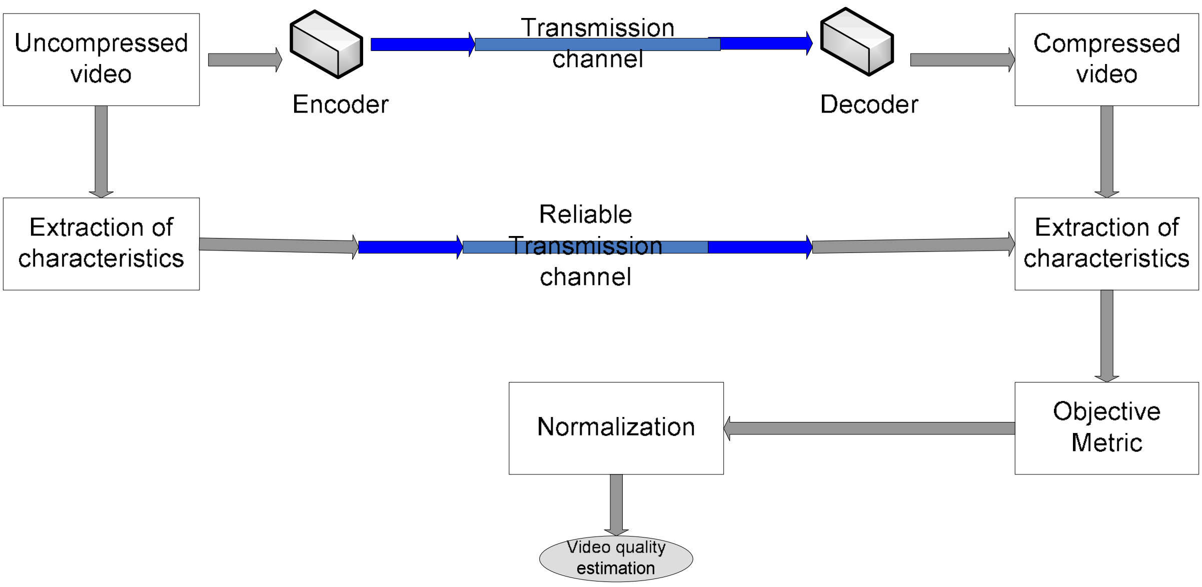


Figure 2.15. Reduced Reference Methods

2.6.5 Non-intrusive Video Quality Measurement

As opposed to both the full reference methods, the no reference methods do not require access to the source video clip. The video quality is predicted entirely from the information provided by the degraded received video clip. In order to monitor the perceptual video quality at the receiver in mobile streaming or IPTV applications, it is difficult to use the full-reference methods as they require access to the source video which may not always be available. This makes no-reference methods attractive to use as they do not require access to the source video clip. Fig. 2.16 gives the conceptual block diagram to predict video quality non-intrusively.

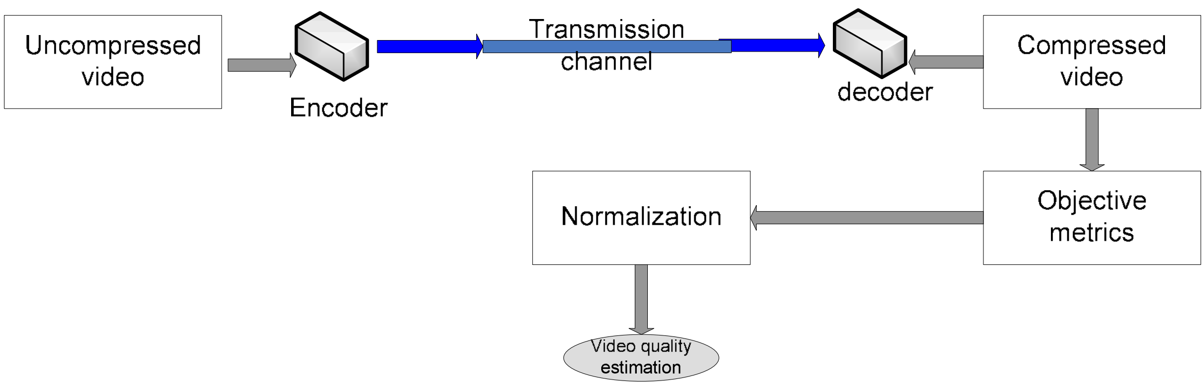


Figure 2.16. Non-intrusive video quality measurement

Non-intrusive video quality measurement can be either (1) Signal-based, (2) Parameter-based or (3) Hybrid (combination of both).

Signal-based non-intrusive video quality measurement

Signal-based non-intrusive video quality measurement measures the input signal only e.g. video content features from raw video.

Parameter-based non-intrusive video quality measurement

Parameter-based non-intrusive video quality prediction can be carried out from a number of parameters either based in application layer (content type, codec, etc.) or from the network (packet/block loss, LBW, delay, etc.) or combining both.

Hybrid non-intrusive video quality measurement

Hybrid non-intrusive video quality measurement combines the signal (bit-stream information) and parameters. The block diagram of hybrid video quality measurement given by ITU-T SG 9 [20] is given in Fig. 2.17.

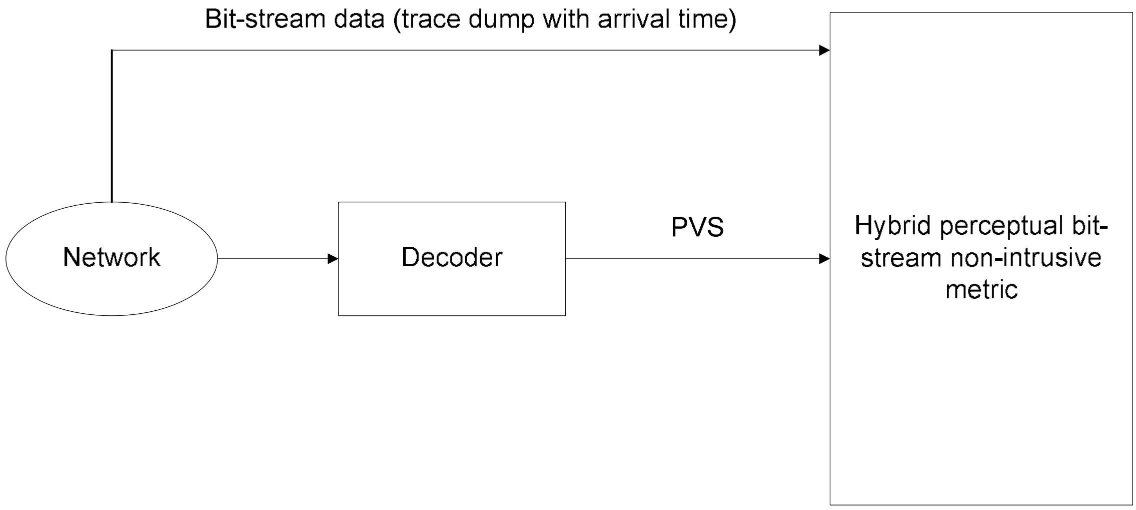


Figure 2.17. Input parameters for non-intrusive hybrid perceptual bit-stream models

The focus of this study is to present non-intrusive video quality prediction models that are based on Qos parameters. The proposed models predict video quality directly from a combination of parameters from the access network, encoder related parameters and content types. The two methods are described briefly.

2.6.5.1 Regression-based methods

In regression-based methods, a number of parameters are used in the regression analysis and model is fitted according to the correlation coefficient that measures the goodness of the fit and the root mean squared error. The block diagram for the regression based and neural network based method is given in Fig. 2.18.

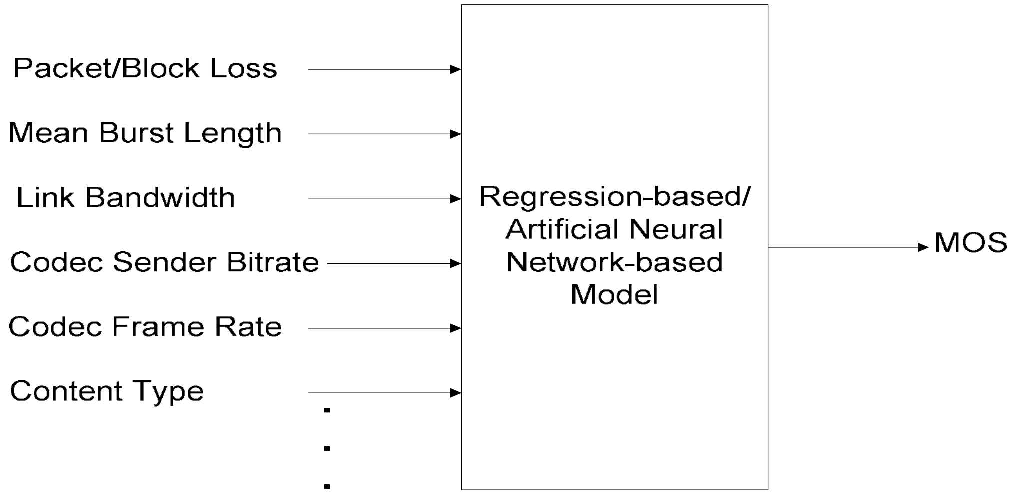


Figure 2.18 Conceptual diagram of Regression/ANN-based model for quality prediction

2.6.5.2 Artificial neural network based methods

Unlike regression-based models that are equation based (mathematical) and are static, ANN model can adjust to the changing environment of WLAN/UMTS access networks because of their capability to learn. ANN models can be trained by learning the nonlinear relationships between perceived video quality (e.g. MOS value) and a combination of application and network level parameters.

Work presented in [14] used ANN to predict video quality over IP networks. The design of ANN model is depicted in Figure 2.18. The ANNs can have a range of input parameters e.g. packet loss, mean burst length, packet size, codec FR, etc. The effect of a combination of access network parameters (e.g. packet/block loss rate, mean burst length and link bandwidth) and encoder related parameters (e.g. codec SBR, FR or CT) on perceived video quality is not very clear. There is a requirement for models to predict video quality to for technical and commercial reasons.

2.7 Summary

The purpose of this chapter has been to present the background on the protocols and architecture of wireless access networks of UMTS and WLAN. The access network performance characteristics are given followed by the video coding concepts and codecs used in the thesis have been introduced. The Quality of Service requirements and the most up to date subjective and objective video quality measurement methods are described in detail. The description of the subjective video quality measurement (e.g. MOS), and objective video quality measurement including both intrusive – full reference (e.g. PSNR and SSIM) and reduced reference and non-intrusive video quality measurement (e.g. regression-based and neural network models) have been presented.

This Chapter has set the background for video quality assessment over wireless access networks and have directed the PhD studies that are presented in Chapters 3 to 7.

Chapter 3

Study of QoS Parameters on video quality over WLAN and UMTS

3.1 Introduction

The main parameters that affect video quality were introduced in section 2.5 (Chapter 2). These impairments include access network impairments of packet/block loss, mean burst length and link bandwidth, encoder impairments of sender bitrate and frame rate and content types. In order to appreciate the relationships between video quality and QoS parameters associated with the encoder, wireless access network and content types, a fundamental investigation was undertaken to quantify their impacts. The metrics used were MOS to measure quality. MOS was obtained by PSNR to MOS conversion from Evalvid [21] (objective method based on simulation as it gave a lot of flexibility). A comprehension of the perceptual effects of these key parameters on video quality is important as it forms the basis for the advancement and development of new and efficient non-intrusive video quality prediction models for video quality monitoring and QoS optimization and control.

Section 3.2 presents the related work on video quality assessment. In Section 3.3 the objective (based on simulation) test set up for WLAN is described. Section 3.4 presents the impact of QoS parameters on video quality over wireless access network of WLAN. The objective test set up over UMTS is presented in Section 3.5. Section 3.6 describes the impact of QoS parameters on quality over wireless access network of UMTS. Section 3.7 gives a detailed discussion of the impact of QoS parameters. Ranking of the QoS parameters are given in Section 3.8. Section 3.9 summarizes the Chapter.

3.2 Related Work

Video quality is impacted by the QoS parameters. This is because low video quality leads to poor QoS as perceived by the user which in turn leads to reduced usage of the applications/services and hence reduced revenue. QoS impacted by the access network parameters can be referred to as NQoS, QoS impacted by the encoder can be referred to as the AQoS and eventually the End-user QoS. End-user QoS is involved with the end-user experience in terms of video quality.

In existing literature, video quality is either predicted using the impact of the NQoS or the AQoS. Rarely parameters associated in both are considered. However, as the end user QoS is impacted by both the AQoS and NQoS, it is important for both content and network providers to understand the capabilities of the network and hence, maximize the QoS of the end user.

Video transmission over UMTS network is subject to quality degradation due to bandwidth limitation when supporting large number of users. This has been addressed by a large number of researchers in [61, 62]-[64]. In [65] it is shown that UMTS radio link of acknowledged mode out performs the unacknowledged mode for video transmission. Similarly, in [66],[67] the transmission requirements and the performance of UMTS dedicated channels of UMTS networks for video streaming has been outlined. In [68] a mechanism for congestion control for video transmission over UMTS networks have been proposed. Whereas, in [69] an error detection scheme is proposed for H.264 encoded videos and in [70] transcoding is used to adapt video content transmitted over UMTS networks. Most of the current work is limited to improving the radio channel. In [71],[72, 73] subjective quality on mobile devices have been evaluated and found combination of bitrates and frame rates for acceptable quality. Work presented in [73] found that H.264 codec produced the most satisfying video quality, however, it was very much content dependent. In [74] H.264 video coding for mobile devices has been assessed and found that the scaling order preferences of the codec are very much content dependent. Works presented in [75] investigate the impact of background traffic on video quality over WLAN and found that the packet size and packet rate of the traffic in the network have a large impact on the video streaming session. Similarly, work presented in [76] concludes with increasing background traffic bandwidth restriction degrades user's Quality of Experience (QoE). Video quality assessment in [77] considers videos in multiple dimensions. They characterize videos in five dimensions such as encoder type, video content, sender bitrate, frame rate and frame size and found that perceptual quality of a decoded video is affected in descending order of significance. In [78] the relation between subjective and objective measures of video quality has been explored. It was found that acceptability depended on more than just visual quality and that both content type and size are important to provide accurate estimates of overall quality in the field of mobile TV. The work in [17] focuses on finding the impact of parameters like sender bitrate, frame rate and packet loss rate on quality using the VQEG subjective database. The authors found that in the application layer sender bitrate and in the network layer packet loss rate have a higher impact on quality. More recently work in [18],[19] presented a study on video quality assessment over wireless networks with H.264 codec. They have made available their subjective database [79] for the research community to use. They found that video quality is impacted both by the encoder and wireless access network impairments. Existing work on video quality assessment is restricted to distortions that occur due to the encoder or the access network. There is very limited research that aims to combine the distortions due to the access network and the encoder. In addition, these distortions are very much content dependent. Hence, the parameters chosen are such that they represent distortions from the encoder, distortions from the access network and represent different types of contents (e.g. slow moving to fast moving). The detailed investigations of the combined parameters from the encoder and access network on different video contents in this Chapter form the basis for the video quality prediction model presented later in this thesis.

3.3 Objective experimental set up over WLAN

The objective experimental set-up for WLAN was based on simulation. The simulation was carried out using NS2 [43] integrated with Evalvid [21] for WLAN. The aim of the set up was to transmit videos with different parameter settings over simulated WLAN network. The raw video was then compared to the degraded video to calculate the end-to-end quality of the received video using PSNR. This allowed a lot of flexibility in the choice of parameters.

3.3.1 Video sequences

Three video clips were chosen as Akiyo (very little movement), Foreman (Some movement) and Stefan (lots of movement). Snapshots of the video clips are given in Fig. 3.1. All sequences can be downloaded from [80] . The video clips were chosen with Quarter Common Intermediate Format (QCIF) (176 x 144). QCIF was specifically chosen as opposed to CIF or larger sizes as it is the recommended size for mobile phones and small hand held terminals which is the target application areas of this study. As the growth in this area has been exponential, the author acknowledges that screen resolution on small handheld devices has increased to 380 x 240 for iphones and even bigger for ipads (ipads were released after this thesis was initially written) and in this sense this work forms the basis for low resolution handheld device and represents worst case scenario. It can easily be extended to higher resolution wireless devices. The results presented here will be valid for higher resolution devices. The models presented later will need to be re-trained.

Figure 3.1. Snap shots of the three video sequences

3.3.2 Data set generation

For quality evaluation a combination of parameters associated with the encoder, WLAN access network and content types were chosen. The parameters associated with the encoder were Frame Rate (FR), and Sender Bitrate (SBR), in the WLAN access network, the parameters chosen were Link Bandwidth (LBW) and Packet Error Rate (PER). Three video sequences were chosen. The video sequences along with the combination parameters chosen are given in Table 3.1.

Video sequences – the video clips chosen from low spatio-temporal activity to high ST activity. In total 3 video clips were chosen as shown in Table 3.1.

The frame rate - the number of frames per second. It takes one of three values as 10, 15 and 30fps.

The sender bitrate - the rate of the encoders output. It is chosen to take 18, 44, 80, 104 and 384kbps.

The link bandwidth: the variable bandwidth link between the routers (Fig. 3.1). It takes the values of 32, 64, 128, 256, 384, 512, 768 and 1000kbps.

Packet Error Rate: the simulator (NS2) drops packet at regular intervals using the random uniform error model, taking five values as 0.01, 0.05, 0.1, 0.15 and 0.2. It is widely accepted that a loss rate higher than 0.2 (20%) will drastically reduce the video quality.

Table 3.1 Dataset combinations over WLAN

Video sequences	Frame Rate (fps)	SBR (kbps)	LBW (kbps)	PER
Akiyo	10, 15, 30	18, 44, 80	32, 64, 128, 256	0.01, 0.05, 0.1, 0.15, 0.2
Foreman				
Stefan		80, 104, 384	384, 512, 768, 1000	

The codec chosen is MPEG4 open source ffmpeg [45] and video sequences were in Quarter Common Intermediate Format (QCIF). QCIF was specifically chosen as the end application was mobile terminal with a small screen. A total of 450 samples were generated using the combinations in Table 3.1.

3.3.3 Experimental set-up over WLAN based on NS2

The experimental set up is given in Fig 3.2. There are two sender nodes as CBR background traffic and MPEG4 video source. Background traffic was added to make the simulation more realistic. Both the links pass traffic at 10Mbps over the Internet which in turn passes the traffic to another router over a variable link. The second router is connected to a wireless access point at 10Mbps and further transmits this traffic to a mobile node at a transmission rate of 11Mbps 802.11b WLAN. The delay is fixed at 1ms. As the aim is to assess the impact of the access network, no packet loss takes place in the wired sector of the video delivery route.

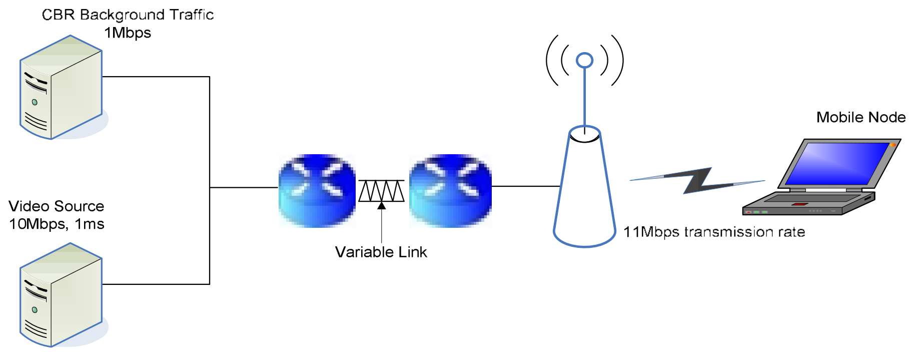


Figure 3.2. Experimental set-up – video transmitted over WLAN

3.3.4 Transmission of video over simulated NS2 network

The transmission of MPEG4 encoded video over WLAN network is illustrated in Fig. 3.3.

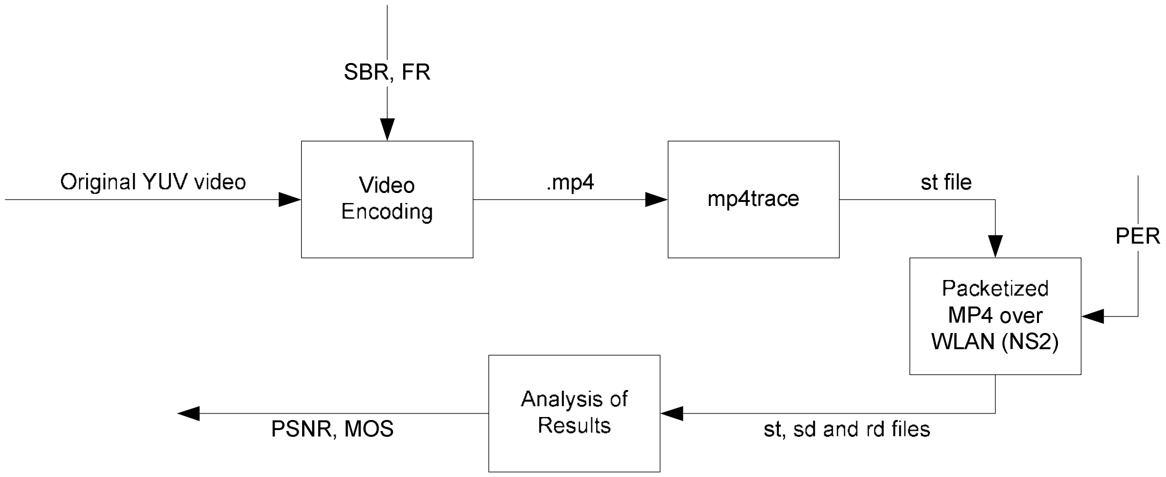


Figure 3.3. Simulation Methodology over NS2 (WLAN)

The original YUV (4:2:0) sequences with QCIF resolution were encoded with MPEG4 with varying values of SBR and FR. The resulting *.m4v video track is then hinted using MP4Box [81] which emulates the streaming of the *.mp4 video over the network based on RTP/UDP/IP protocol stack. The maximum transmission packet size is 1024 bytes. The resulting trace file feeds the NS2 simulation model as required. Random uniform error model is used to deliver the video packets initially. The link bandwidth in Fig. 3.3 is varied from 32kbps to 1Mbps according to Table 3.1. This is done to measure the impact of link bandwidth on end-to-end quality. The incoming trace file (st), the sender module (sd) and receiver module (rd) are used by the etmp4 [21] program to generate *.mp4 file. Finally, the degraded video is generated using ffmpeg decoder. The PSNR is then computed by comparing the original video to the degraded video. The CBR rate is fixed to 1Mbps to simulate realistic scenario. Further, the experiment is repeated to account for bursty errors. The average PER is then calculated from Eq. (2.1) and Eq. (2.2). In this thesis the values of P_{GG}, P_{BB} and P_G are set to 0.96, 0.94 and 0.001 respectively from [60]. The value of P_B is set from 0.05 to 0.5 with 0.05 intervals to correspond with packet error rates of 0.02, 0.04 to 0.2 in accordance with the formula given by Eq. (2.2).

To account for different packet loss patterns, 10 different initial seeds for random number generation were chosen for each packet error rate. All results generated in the thesis were obtained by averaging over these 10 runs.

All the experiments in this thesis were conducted with an open source framework Evalvid [21] and network simulator tool NS2 [43]. Video quality is measured by taking the average PSNR over all the decoded frames. MOS scores are calculated based on the PSNR to MOS conversion from Evalvid [21]. The mapping is given in Table 3.2.

Table 3.2 PSNR to MOS conversion

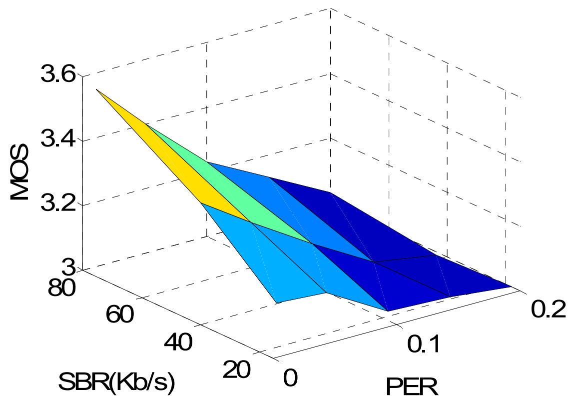
PSNR (dB)	MOS
>37	5
31 – 36.9	4
25 – 30.9	3
20 – 24.9	2
< 19.9	1

3.4 Impact of QoS parameters over WLAN

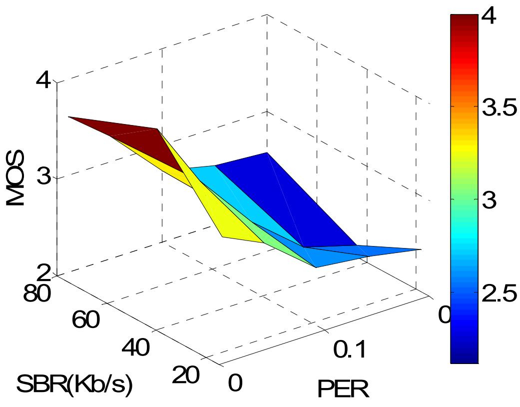
In this section the effects of the four parameters on video quality is presented. Three-dimensional figures were chosen in which two parameters were varied while keeping the other two fixed. MOS is used as a metric to measure quality.

3.4.1 Impact of SBR and PER on quality

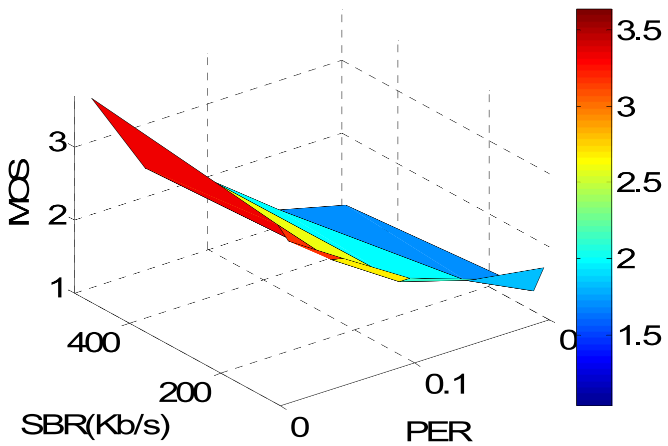
To evaluate the impact of SBR and PER on quality (MOS), the frame rate was kept fixed at 10fps and the link bandwidth was fixed at 128kb/s. Fig. 3.4a shows the MOS values for ‘Akiyo’. It was observed that the MOS dropped to 3 when the packet loss was 20% which is an acceptable value for communication quality. This shows that when there is very little activity in content the video quality is still acceptable at low send bitrates and with high packet loss.



(a) MOS vs SBR vs PER for 'Akiyo'



(b) MOS vs SBR vs PER for 'Foreman'



(c) MOS vs SBR vs PER for 'Stefan'

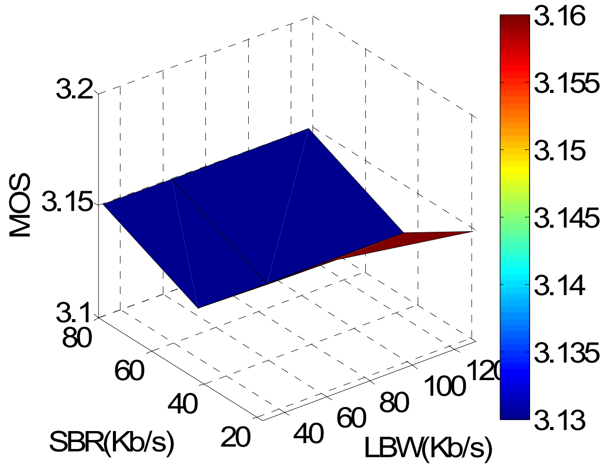
Figure 3.4. MOS vs SBR vs PER

Fig. 3.4b show the MOS scores for 'Foreman'. The frame rate is fixed at 10fps and the link bandwidth at 384kb/s. It was observed that with higher sender bitrate of 80kb/s the video quality is very good (MOS > 3.5), however, the quality fades rapidly with increasing packet loss.

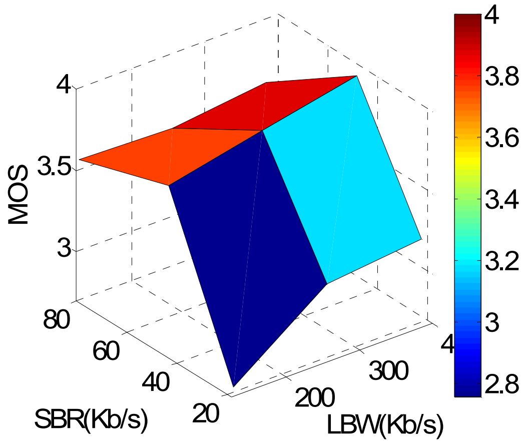
Fig.3.4c show the MOS scores for ‘Stefan’. The frame rate was kept fixed at 10fps and the link bandwidth was fixed at 512kb/s. Again, the video quality is very good for higher sender bitrate of 512kb/s, but fades very rapidly with increasing packet loss.

3.4.2 Impact of SBR and LBW on quality

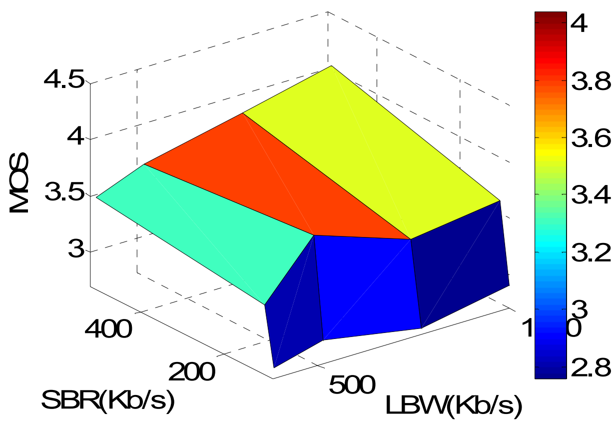
To evaluate the impact of SBR and LBW on quality, the frame rate is fixed at 10fps without any packet loss for three video sequences. In Figs 3.5a, b and c increasing the link bandwidth only improves the MOS score if the video is encoded at a bitrate less than the LBW. However, a worsening in video quality was observed if the sender bitrate is more than the link bandwidth due to congestion.



(a) MOS vs SBR vs LBW for Akiyo



(b) MOS vs SBR vs LBW for Foreman

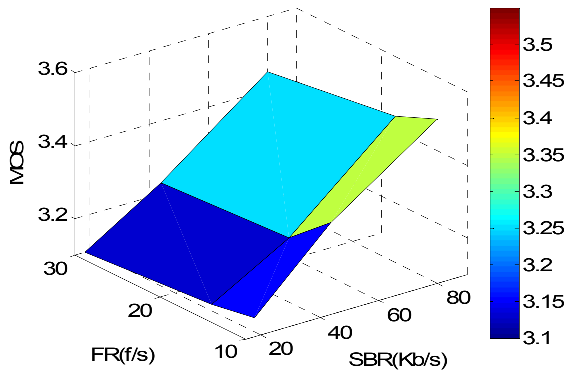


(c) MOS vs SBR vs LBW for Stefan

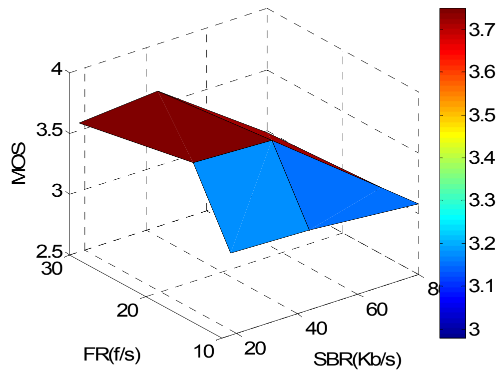
Figure 3.5. MOS vs SBR vs LBW for the three video sequences

3.4.3 Impact of SBR and FR on quality

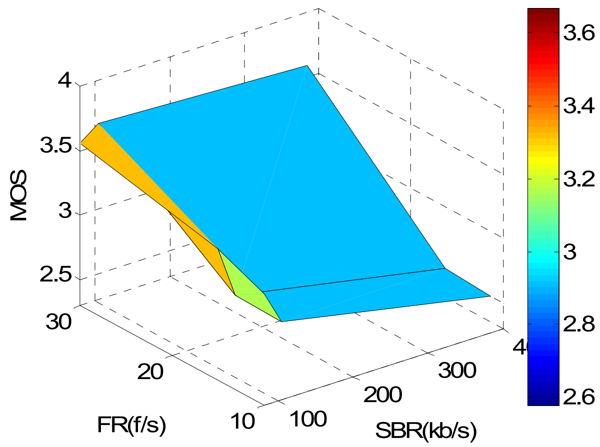
To evaluate the impact of SBR and FR on quality, packet loss is zero and LBW is fixed at 256kbps for video sequences of 'Akiyo' and 'Foreman' and 512kbps for 'Stefan'. From Figs. 3.6a,b&c it was found that frame rate is not as significant as sender bitrate. Improvement in video quality is only achived upto frame rates of 15fps. This confirms that for low sender bitrate videos low frame rates give better quality (e.g. frame rate ≤ 10fps). However, for higher sender bitrate higher frame rate will not reduce video quality.



(a) MOS vs FR vs SBR for Akiyo



(b) MOS vs FR vs SBR for Foreman



(c) MOS vs FR vs SBR for Stefan

Figure 3.6. MOS vs FR vs SBR for the three video sequences

3.5 Objective experimental set up over UMTS

The objective experimental set-up over UMTS was based on simulation. The simulation was carried out using NS2 [43] and integrated with Evalvid [21] and Eurane [53] for UMTS. Also OPNET simulation platform was used for the UMTS in addition to NS2. Video sequences chosen are the same as described in sub-section 3.5.1 over WLAN. OPNET is another simulation platform for UMTS access network. The datasets generated from using both simulation networks were compared and analyzed.

3.5.2 Data set generation

Blocks are transmitted over the physical channel using the UMTS physical model. The channel bitrates and the Transmission Time Interval (TTI) affiliated with the channel that are used in the simulation are shown in Table 3.3. The UMTS downlink bitrate has one of three values as 128, 256 and 384kbps. As the physical layer passes the block to the Medium Access Layer (MAC) including the error indication from the Cyclic Redundancy Check (CRC), the Block Error Rate (BLER) is then defined as the overall probability of the block error from the output of the Physical layer in this thesis. Therefore, an error model based on both uniform distribution of block errors and 2-state Markov model with variable Mean Burst Lengths (MBL) was used in the simulation. The BLER ranges between 0-20% in the simulation. Three values of MBL were selected based on the mean error burst length found in [82] from real-world UMTS measurements. MBL was chosen as 1 (to depict random uniform losses), 1.75 (typical of mobile roaming) and 2.5 (very bursty situations). MBL is calculated according to sub-section 2.3.2 in Chapter 2. UMTS access network parameters chosen were BLER and MBL. Encoder parameters were FR and SBR. Frame rate ranges from 5fps to 15fps, whereas, the sender bitrate ranges from 18kbps to 104kbps for the three video sequences chosen. See Table 3.3 for details.

Table 3.3 Dataset combination over UMTS

Video sequences	FR (fps)	SBR(kbps)	LBW	BLER (%)	MBL
Akiyo, Foreman	5, 7.5,10, 15	48, 88,128	128, 256, 384	1, 5, 10, 15,20	1, 1.75, 2.5
Stefan		88,130,256			

Table 3.4 Encoding Parameter Set-up

```
InputFile             = "foreman_qcif.yuv"       # Input sequence
InputHeaderLength     = 0      # If the inputfile has a header, state it's length in byte her
StartFrame            = 0      # Start frame for encoding. (0-N)
FramesToBeEncoded     = 300      # Number of frames to be coded
FrameRate             = 30   # Frame Rate per second (0.1-100.0)
SourceWidth           = 176    # Source frame width
SourceHeight          = 144    # Source frame height
SourceResize          = 0      # Resize source size for output
OutputWidth           = 176    # Output frame width
OutputHeight          = 144    # Output frame height

TraceFile             = "foremanstats/foreman_encl548.txt"      # Trace file
ReconFile             = "foremanstats/foreman_recl548.yuv"       # Recontruction YUV file
OutputFile            = "foremanall/foreman_qcif1548.264"         # Bitstream
StatsFile             = "foremanstats/statsforeman1548.dat"         # Coding statistics file

##########################################################################################
# Encoder Control
##########################################################################################
ProfileIDC            = 66 # Profile IDC (66=baseline, 77=main, 88=extended; FREXT Profiles:
100=High, 110=High 10, 122=High 4:2:2, 244=High 4:4:4, 44=CAVLC 4:4:4 Intra)
IntraProfile          = 0   # Activate Intra Profile for FRExt (0: false, 1: true)
                            # (e.g. ProfileIDC=110, IntraProfile=1  =>  High 10 Intra Profile
LevelIDC              = 12  # Level IDC   (e.g. 20 = level 2.0)
# In Baseline, Main and Extended: Level 1b is specified with LevelIdc==11 and
constrained_set3_flag == 1

IntraPeriod           = 0   # Period of I-pictures   (0=only first)
IDRPeriod             = 0   # Period of IDR pictures (0=only first)
AdaptiveIntraPeriod   = 1   # Adaptive intra period
AdaptiveIDRPeriod     = 1   # Adaptive IDR period
IntraDelay            = 0   # Intra (IDR) picture delay (i.e. coding structure of PPIPPP... )
EnableIDRGOP          = 0   # Support for IDR closed GOPs (0: disabled, 1: enabled)
EnableOpenGOP         = 0   # Support for open GOPs (0: disabled, 1: enabled)
QPISlice              = 28  # Quant. param for I Slices (0-51)
QPPSlice              = 28  # Quant. param for P Slices (0-51)
FrameSkip             = 1   # Number of frames to be skipped in input (e.g 2 will code every
third frame)
ChromaQPOffset        = 0   # Chroma QP offset (-51..51)
```

The video sequences along with the combination of parameters chosen are given in Table 3.3. As the transmission of video was for mobile handsets, all the video sequences are encoded with H.264/AVC codec, with Baseline Profile at 1.2 level, and with a QCIF resolution using the Joint Model (JM) reference software [46] developed by JVT for testing. The JM software comprises of both the encoder and the decoder that are compliant with the H.264/AVC . The encoder/decoder settings are passed via command line and/or the configuration files encoder.cfg and decoder.cfg. The considered frame structure is IPPP for all the sequences, since the extensive use of I frames could saturate the available data channel. Further information about the encoding process can be found in [58]. From the 3GPP recommendations for video streaming services, such as VOD or unicast IPTV services, a client should support H.264 (AVC) Baseline Profile up to the Level 1.2. [46]. From these considerations, the encoding features were set. A snapshot of the encoder.cfg for 'Akiyo' sequence is shown in Table 3.4. Framskip defines the frame rate. Frameskip=3 means a frame rate of 10fps.

3.5.3 Experimental set-up over UMTS based on NS2

The network topology is modelled in the UMTS extension for the NS2 [43] namely, Enhanced UMTS Radio Access Network Extension (EURANE) [53] integrated with Evalvid [21] for H.264 video streaming. NS2 was chosen due to its flexibility and based on the characteristics of the link bandwidth. H.264 codec is chosen as opposed to MPEG4 as it is the recommended codec for low bitrate transmission.

The simulation model is given in Fig. 3.7. It consists of a streaming client and server. In the simulation set up given in Fig. 3.7, the streaming client is given by the User Equipment (UE) and the streaming server which is located in the Internet is a fixed host. The addressed scenario comprises of a UMTS radio cell covered by a node B connected to an RNC. The simulation model consists of a UE connected to Downlink Dedicated Physical Channel (DPCH).

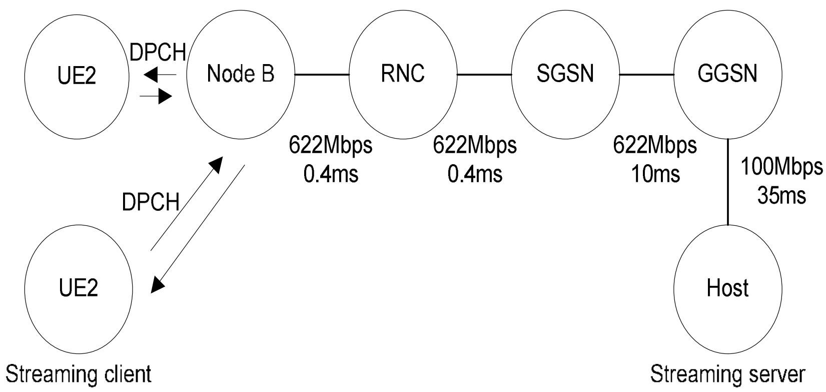


Figure 3.7. UMTS Network Topology

The aim of the thesis was to investigate the impact of UMTS access network losses, the simulation was set up in such a way that only the access network was modelled for streaming H.264 video. Hence, there were no packet losses on either of the Internet or the UMTS core network (e.g. SGSN, GGSN). The function of SGSN and GGSN were then modelled as traditional ns nodes as were wired. In Fig. 3.7 the links between the two nodes are labelled with their bitrate (in bits per second) and delay (in mili seconds). The link capacity is chosen such that the radio channel is the bottleneck in the connection. At present, in the Packet Data Convergence Protocol (PDCP) layer no header compression technique is supported. The simulation parameters are summarized in Table 3.5.

Table 3.5 Simulation Parameters for UMTS

Input parameter	Value
UMTS physical channel type	DPCH
Downlink bitrate (kbps)	128, 256, 384
Uplink bitrate(kbps)	64
Downlink TT1 (ms)	10
Uplink TT1 (ms)	20
Packet size (bytes)	1024
UDP header size (bytes)	8
IP header size (bytes)	20
RLC mode	Acknowledged
Encoder Profile/Level IDC	(66,11), Baseline Profile, level 1.2
Sequence type	IPPP
Entropy coding method	CAVLC
BLER	0 -20%
Error model	2-state Markov

3.5.4 Experimental set-up over UMTS based on OPNET

UMTS access network was further simulated over OPNET. OPNET was used in addition to NS2 to analyze the specific impact of the UMTS error conditions into the perceived video quality, due to its accuracy of the implementation for the RLC not-in-order delivery mechanism. The UMTS network topology modelled in OPNET Modeler® [44] is shown in

Fig. 3.8. It is made up of a Video Server, connected through an IP connection to the UMTS network, which serves to the mobile user.

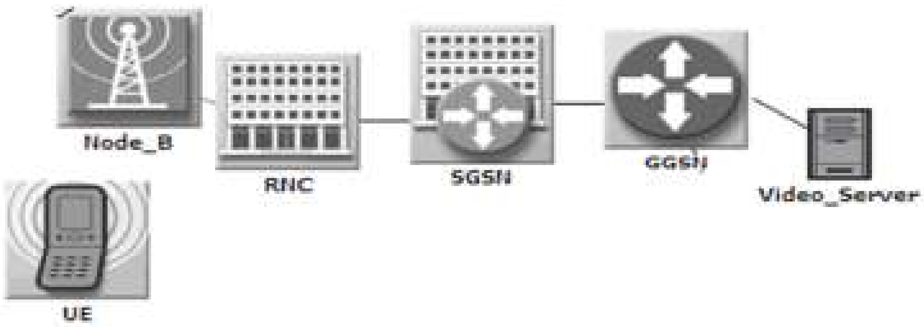


Figure 3.8. OPNET network scenario

With regard to the UMTS arrangement, the video transmission is sustained over a Background Packet Data Protocol (PDP) Context with a typical mobile wide area configuration as defined in 3GPP TR 25.993 [48] for the "Interactive or Background / UL:64 DL:384 kbps / PS RAB". The transmission channel supports maximum bitrates of 384 kbps Downlink / 64 kbps Uplink over a Dedicated Channel (DCH). Since the analyzed video transmission is unidirectional, the uplink characteristics are not considered a bottleneck in this case. The parameter set up is the same as Table 3.5.

The UMTS link layer model that is actualized is based on the results presented in [82], which analyzes the error traces from currently deployed 3G UMTS connections. Specifically, the error model at RLC layer indicates that, for mobile users, the channel errors can be accumulated at Transmission Time Interval (TTI)-level. This error model leads to possible losses of RLC SDUs, which lead to losses at RTP layer, and finally to frame losses at video layer.

3.5.5 Transmission of video over simulated (NS2/OPNET) network

The transmission of H.264 encoded video over UMTS network is illustrated in Fig. 3.9. The original YUV sequences are encoded with the H.264/AVC JM Reference Software with varying SBR and FR values as shown in Table 3.3. H.264 is chosen as it is the recommended codec to achieve suitable quality for low sender bitrates. The resulting *.264 video track becomes the input of the next step, which emulates the streaming of the mp4 video over the network based on the RTP/UDP/IP protocol stack. The maximum packet size is set to 1024 bytes in this case. The resulting trace file feeds the OPNET/NS2 simulation model as required. The incoming trace file (st), the sender module (sd) and the receiver module (rd) are used by the etmp4 program to generate *.264 files. Finally, the last step is in charge of analyzing the quality of the received video sequences against the original quality and the resulting PSNR values are calculated with the ldecod tool included in the H.264/AVC JM Reference Software. MOS scores are calculated based on the PSNR to MOS conversion from Evalvid [21] and as shown in Table 3.2.

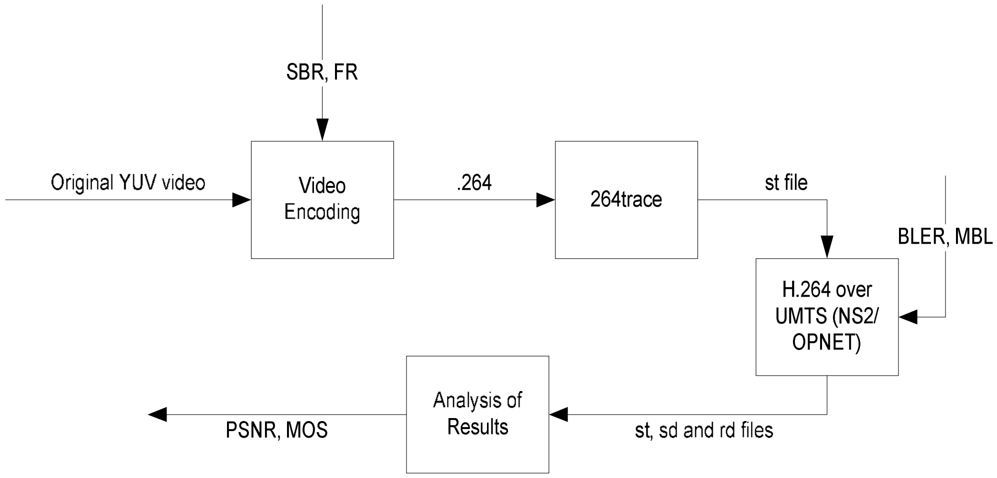


Figure 3.9. Simulation methodology over NS2/OPNET (UMTS)

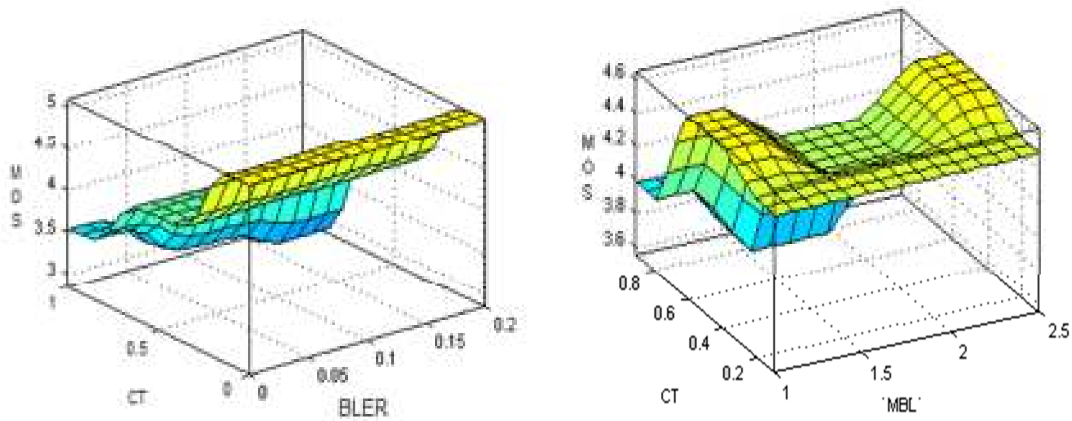
Instead of setting up a target BLER value for the PDP Context, the UE model is modified in order to support the desired error characteristics. The implemented link loss model is a 2-state Markov model and its performance is provided by two parameters: BLER and the MBL.

3.6 Impact of QoS parameters over UMTS

The next two sub-sections describe the joint impact of encoder and UMTS access network parameters for three different video sequences on quality. The video sequences are classified in Chapter 4 as Content Types (CT). However, here the three video sequences take three discrete values as CT=0.1 (Akiyo), CT=0.5 (Foreman) and CT=0.9 (Stefan)... The results presented in this section take the data generated using OPNET. Very little difference was found between the data generated from NS2 and OPNET. A detailed analysis is presented in Section 3.7.

3.6.1 Impact of BLER and MBL on Content Type (CT)

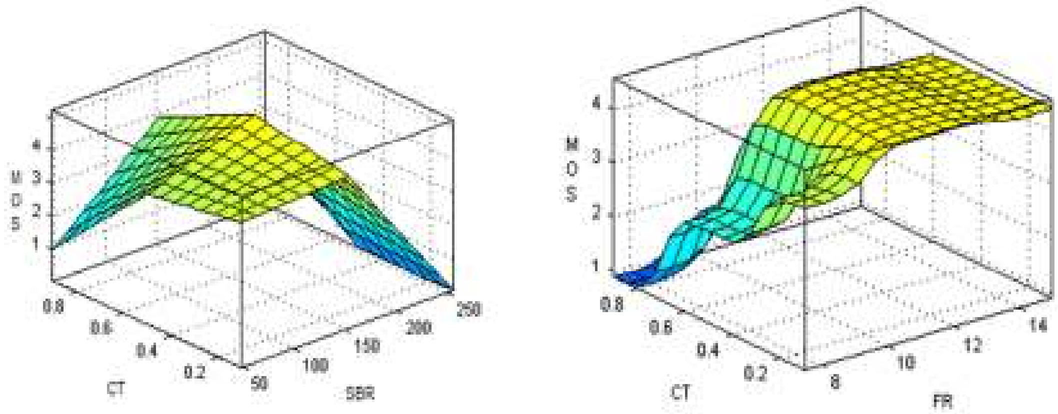
The impact of MBL and BLER on our chosen content types are given in Figs. 3.10a and b.



(a) MOS vs CT vs BLER (b) MOS vs CT vs MBL

Figure 3.10. MOS vs CT vs BLER

Three video sequences were chosen and are defined as Akiyo (CT=0.1), Foreman (CT=0.5) and Stefan (CT=0.9) from slow moving to fast moving sports type of content. This is further verified with the content classification given in Chapter 4. From Fig. 3.10a it is observed that as the activity of the content increases the impact of BLER is much higher. For example, for 20% BLER, CT of slow to medium type gives very good MOS, whereas as the content activity increases, MOS reduces to 3. From Fig. 3.10b it was observed that the MBL similar to BLER has greater impact for content types with higher S-T activity.



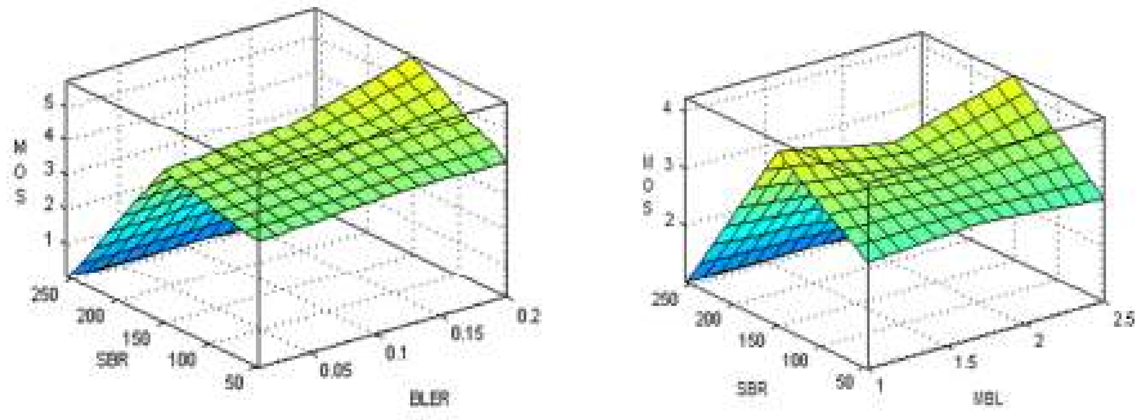
(a) MOS vs CT vs SBR (b) MOS vs CT vs FR

Figure 3.11. MOS vs CT vs SBR

Similarly, the impact of SBR and FR on CT is given by Figs. 3.11a and b. Again it is observed that as the activity of content increases for very low SBRs (20kb/s) and low FRs (5f/s) the MOS is very low. However, for slow to medium content activity the impact of SBR and FR is less obvious. The lower value of MOS for higher SBR is due to network congestion.

3.6.2 Impact of BLER and MBL on SBR

The combined impact of SBR and BLER/MBL is given in Fig. 3.12a and b. As expected, with increasing BLER and MBL the quality reduces. However, for increasing SBR the quality improves up to a point (SBR ~ 80kb/s) then increasing the SBR results in a bigger drop of quality due to network congestion. From Fig. 3.12b it is observed that the best quality in terms of MOS was for an MBL of 1 (depicted random uniform scenario). This would be expected because the BLER was predictable. The worst quality was for BLER of 2.5 (very bursty scenario). Again this substantiates previous findings on 2-state Markov model. It was interesting to observe how MBLs impact on quality, however it is captured by the QoS parameters of BLER. Similar to Fig. 3.12a for high SBR, quality collapse for all values of MBLs due to network congestion.

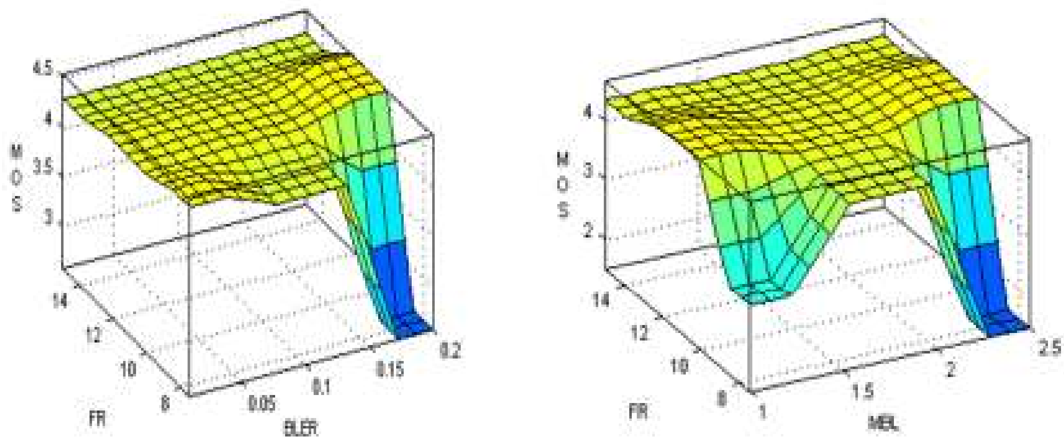


(a) MOS vs SBR vs BLER (b) MOS vs SBR vs MBL

Figure 3.12. MOS vs SBR vs BLER

3.6.3 Impact of BLER and MBL on FR

Figs. 3.13a and b show the impact of BLER and MBL on FR for all content types. It is observed that for faster moving contents very low frame rates of 7.5 fps impair quality. Again, it is observed that both BLER and MBL impact on the overall quality. The impact of frame rate is more obvious for low FRs and high BLER. However, when BLER is low quality is still acceptable. This is shown in Fig. 3.13a. Fig. 3.13b shows that for low FRs quality is acceptable for MBL of 1.75. However, for MBL of 1 it starts to deteriorate. This is mainly for high spatio-temporal contents. However, quality completely collapses for MBL of 2.5 (very bursty scenario). Again the impact is much greater on contents with high spatio-temporal activity compared to those with low ST activity.



(a) MOS vs FR vs BLER (b) MOS vs FR vs MBL

Figure 3.13. MOS vs FR vs BLER/MBL

3.6.4 Impact of LBW on CT

The impact of physical layer parameters of LBW and BLER vary depending on the type of content. For slow moving content BLER of 10% gives acceptable quality, however for fast moving content for the same BLER the quality is completely unacceptable. From Fig. 3.14 it is observed that if the LBW is 128kbps then quality is low due to network congestion (for content encoded at SBR close to 128kbps). The impact of LBW is normally measured by BLER in real systems.

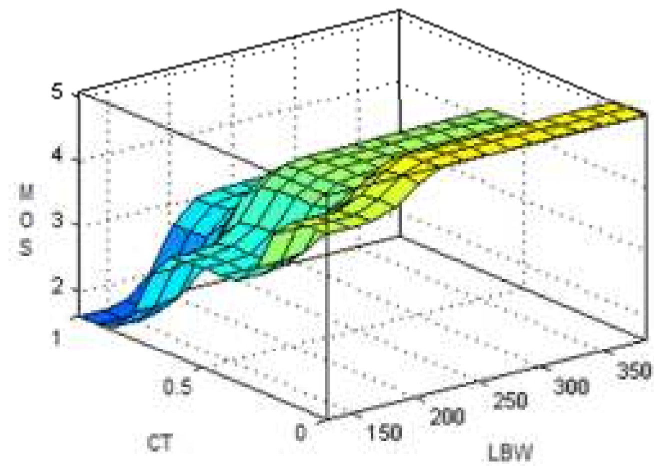


Figure 3.14. MOS vs CT vs LBW

3.7 Discussions and analysis

In order to completely study the influence of different QoS parameters on MOS ANOVA (analysis of variance) [83] was performed on the MOS data set. Three data sets of MOS were generated as follows.

1. Simulation using NS2 over WLAN – 5 parameters CT, SBR, FR, PER and LBW
2. Simulation using NS2 over UMTS – 5 parameters CT, SBR, FR, BLER and LBW
3. Simulation using OPNET over UMTS – 5 parameters CT, SBR, FR, BLER and MBL

Tables 3.6, 3.7 and 3.8 show the results of the ANOVA analysis on three datasets respectively.

Table 3.6 Five-way ANOVA on MOS obtained via NS2 Simulation over WLAN

Parameters	Sum of Squares	Degrees of Freedom	Mean Squares	F statistic	p-value
CT	44.962	2	22.4811	138.38	0
FR	1.887	2	0.9434	5.81	0.0032
SBR	4.569	4	1.1423	7.03	0
PER	97.237	4	24.3093	149.63	0
LBW	22.845	7	3.2635	20.09	0

Table 3.7 Five-way ANOVA on MOS obtained via NS2 Simulation over UMTS

Parameters	Sum of Squares	Degrees of Freedom	Mean Squares	F statistic	p-value
CT	15.052	2	15.052	362.71	0
FR	0.2598	2	0.1299	3.13	0.0489
SBR	4.462	2	2.23098	3.99	0.0217
BLER	0.9926	4	0.2481	5.98	0.0003
LBW	35.381	2	17.6905	79.43	0

Table 3.8 Five-way ANOVA on MOS Obtained via OPNET Simulation over UMTS

Parameters	Sum of Squares	Degrees of Freedom	Mean Squares	F -statistic	p-value
CT	29.508	2	14.754	109.27	0
FR	1.017	2	1.016	7.53	0.0069
SBR	9.559	2	4.7797	35.4	0
BLER	1.152	4	0.3839	2.84	0.0402
MBL	0.361	2	0.1807	1.34	0.2659

5-way ANOVA (Tables 3.6, 3.7 and 3.8) was performed on the MOS datasets to determine if the means in the MOS in all three data sets given by the QoS parameters differ when grouped by multiple factors (i.e. the impact of all five parameters on MOS). Tables 3.6, 3.7 and 3.8 show the results, where the Sum of Squares is represented in the first column, second column is the Degrees of Freedom, the third column is the Mean Squares which is defined as the ratio of Sum of Squares to Degrees of Freedom. The F statistic is shown in the fourth column and the p-value is given in the fifth column. The p-value is determined from the cumulative distribution function (cdf) of F [83]. A small p-value ($p \leq 0.01$) indicates that the MOS is considerably affected by a variation of the corresponding parameter. Besides, from the magnitudes of p-values (in all three tables), it is observed that CT (p-value=0 in all three datasets), SBR (p-value=0 in 2 datasets), LBW (p-value=0 in two datasets) and PER (p-value=0) impacts the MOS results the most, followed by BLER and FR. While, MBL has the least influence on quality. As the MOS is found to be mostly affected by CT, SBR and LBW, it can be further categorized the CT, SBR and LBW using the Multiple Comparison Test (MCT) based on Tukey-Kramer's Honestly Significant Difference (HSD) criterion [84]. The results of comparison test for CT, SBR and LBW are shown in Figs. 3.15a, b and c, the mean and the 95% confidence interval are indicated by the centre and span of each horizontal bar respectively.

The studies numerically substantiate the following observations of video quality assessment as:

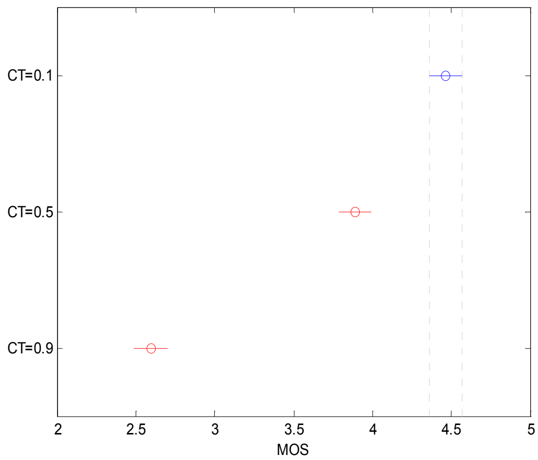
- The most important QoS parameter in the application layer is the content type. Therefore, an accurate video quality prediction model must consider all content types. The values of FR and SBR are independent from the content type and therefore cannot provide accurate estimation of quality.

- The optimum combination of SBR and FR that achieves the best quality is very much content dependent and varies from sequence to sequence. It was found that for slow moving content FR=5 and SBR=18kbps gave acceptable quality, however as the spatio-temporal activity of the content increased this combination gave unacceptable quality under no network impairment. This is shown by Figs. 3.11a and b. From Fig. 3.11a when CT=0.9(Stefan), SBR=20kps for a FR of 10fps MOS=1. Similarly, from Fig. 3.11b when FR=7.5fps for content type of Stefan at an SBR of 48kps MOS=1. However, when CT=0.5/0.1 (Foreman/Akiyo) then for the same conditions MOS increases to 2.5. This clearly shows that as the spatio-temporal activity of the content increases low FRs and SBRs give very low quality. However, with slow to medium spatio-temporal activity the low FR-SBR combination gives acceptable quality. Hence the choice of SBR and FR are very much dependent on the type of content. Also knowing the initial encoding SBR saves useful bandwidth resources.
- The ANOVA results showed that the QoS parameter of LBW had a significant impact on quality. In real systems the impact of LBW is generally measured by BLER. Therefore, an accurate video quality prediction model must take into account the influence of physical layer in addition to application layer parameters.
- The most important QoS parameter in the physical layer is BLER. Therefore, an accurate video quality prediction model must consider the impact of physical layer in addition to application layer parameters. The impact of physical layer parameters of LBW and BLER vary depending on the type of content. For slow moving content BLER of 10% gives acceptable quality, however for fast moving content for the same BLER the quality is completely unacceptable. Therefore, the impact of physical layer QoS parameters is very much content dependent as well. This is

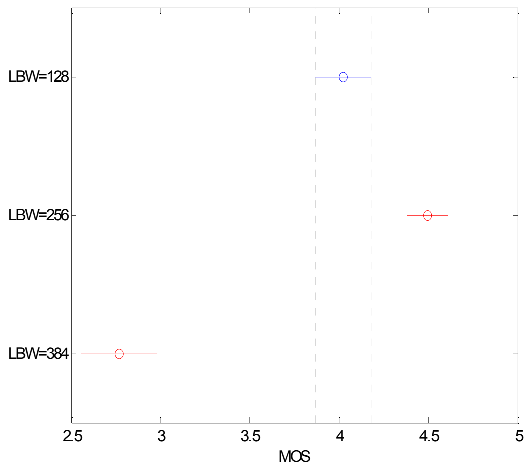
explained in Figs. 3.11a and 3.14. From Fig. 3.11a it is observed that as the activity of the content increases the impact of BLER is much higher. For example, for 20% BLER, CT of slow to medium type gives very good MOS, whereas as the content activity increases, MOS reduces to 3. From Fig. 3.14 it is observed that if the LBW is 128kbps then quality is low due to network congestion (for content encoded at SBR close to 128kbps). The impact of LBW is normally measured by BLER in real systems.

- The impact of physical layer parameters of MBL and BLER vary depending on the type of content. For slow moving content BLER of 20% gives acceptable quality, however for fast moving content for the same BLER the quality is completely unacceptable. Therefore, the impact of physical layer QoS parameters is very much content dependent.
- Finally, comparing the results of Tables 3.7 (over NS2) and Table 3.8 (over OPNET), CT was found to be the most important parameter (P-value=0) from both datasets. FR had similar impact from both datasets. The difference was found in BLER and SBR. The dataset generated over NS2 found BLER to be more important than SBR, whereas, the dataset generated over OPNET found SBR to be more important than BLER. This was due to the fact the dataset generated over OPNET took maximum SBR of 256kbps. This SBR gave very low MOS values due to congestion (especially at LBW of 256kbps). The maximum SBR taken over NS2 was 128kbps and hence, did not result from congestion. However, this shows that (i) the difference in datasets over the two simulation environments was negligible and (ii) it helped in deciding the ranking of QoS parameters in order of importance (see next section). The order of SBR and BLER very much depends on network conditions e.g. if a network is congested then BLER is more important, as low SBR

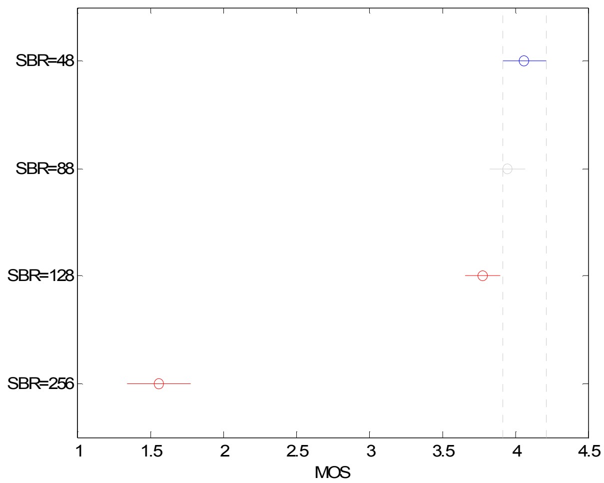
will give acceptable quality, however, if the network conditions are good then the same SBR gives unacceptable quality making SBR more important than BLER.



(a) CT vs MOS



(b) LBW vs MOS



(c) SBR vs MOS

Figure 3.15. Multiple comparison test for MOS

3.8 Ranking of QoS parameters over WLAN and UMTS

The QoS parameters chosen for the model development (see Chapters 5 and 6) can be summarized in Fig. 3.16. Following analysis from previous section the QoS parameters have been ranked as shown in Table 3.10. Content types are ranked in the highest order as the impact of PER/BLER is very much content dependent. As an example, if network providers assure a maximum of 5% PER/BLER then content type of news will have an acceptable quality (MOS ~ 3.5), however, for sports that quality will drop to around MOS of 2. It was found that FR and MBL have the least impact on quality. PER/BLER/LBW comes second in the ranking order. LBW is generally measured in terms of packet/blocks lost. SBR came third in the ranking order. However, it can also be regarded as a joint second with PER/BLER/LBW. SBR of 18kbps gave acceptable quality for slow moving content. However, the same SBR gave unacceptable quality as the ST-activity of the content increased under no network losses.

Table 3.9 QoS parameters ranking order

QoS Parameters	Ranking order
Content type	1
PER/BLER/LBW	2
SBR	3
FR	4
MBL	5

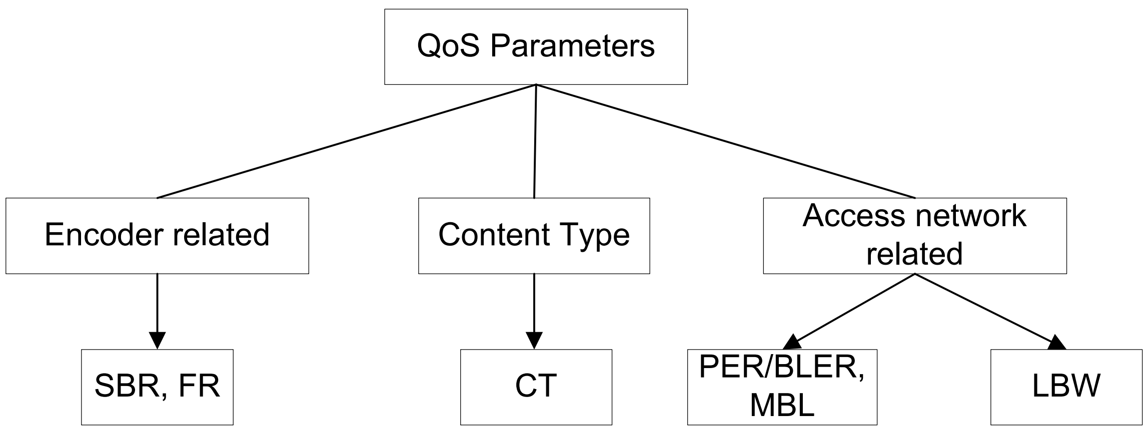


Figure 3.16. QoS Parameters

3.9 Summary

A crucial analysis between video quality and access network impairments (e.g. packet/block loss, LBW and MBL) and application level impairments (CT, SBR and FR) have been undertaken using datasets generated by objective measurement of PSNR and MOS obtained from PSNR to MOS conversion over WLAN and UMTS. The results show that in the application layer CT and SBR have a greater impact, whereas in the physical layer BLER and LBW have a greater impact. Finally, a ranking order of QoS parameters has been established in order of importance. The results from this Chapter enables to decide on the choice of parameters required for developing statistical and neural network models for video quality prediction, non-intrusively which are presented in Chapters 5 and 6 respectively.

Content type was found to be the most important parameter and was highest in the ranking order. Therefore, the next Chapter presents methods for classifying contents.

Chapter 4

Content classification Methods

4.1 Introduction

In Chapter 3 an investigation on the impact of QoS parameters on video quality was presented. The QoS parameters were further ranked in order of importance. Content type was found to be the most important QoS parameter. Therefore, the aim of this Chapter is to present methods to classify video contents objectively which is then used as an input to the non-intrusive prediction model (See Chapters 5 and 6).

Section 4.2 presents the related work on content classification. Video content types are listed in Section 4.3. Content dynamics are discussed in Section 4.4. Methods to classify the video contents based on ST-feature extraction are presented in Section 4.5. Section 4.6 presents the alternate method for content classification. Section 4.7 summarizes the Chapter.

4.2 Related work

In Chapter 3 it was concluded that video content has an impact on video quality achieve able under same network condition. Recent work has also shown the importance of video content in predicting video quality. Feature extraction is the most commonly used method to classify video contents. In [11], [85],[86],[87] video content is classified based on the spatial (edges, colours, etc) and temporal (movement, direction, etc) feature extraction which were then used to predict video quality together with other application-level parameters such as sender bitrate and frame rate. The limitations of using feature extraction is that it does not express the semantic scene importance but the formal features such as degree of motion compensation. Work presented in [88] has explored the impact of spatial and temporal features on video quality prediction. Recent studies in [89],[90] have classified video content based on content characteristics obtained from users' subjective evaluation using cluster [91] and Principal Component Analysis (PCA) [83]. In [92],[93] a combination of PCA [83] and feature extraction to classify video contents have been presented.

The work presented in Chapter 3 concluded that video quality is affected by the distortions caused by the access network and the encoder. However, the impact of these distortions is very much content dependent. Hence, for a video quality prediction model it is important to consider the impact of different types of content.

4.3 Video clips/sequences

The chosen video sequences ranged from very little movement, i.e. with a fixed background, small moving region of interest to fast moving sports clips. The choice of video sequences was to reflect the varying spatio-temporal activity of the content representative of typical content offered by content providers e.g. news type of content or fast moving sports content.

Twelve video sequences were used that ranged from slow movement (head and shoulder) to fast moving sports clips. The snap shot of the video sequences are shown in Fig. 4.1. All sequences can be downloaded from [80] . The description of the content types is given in Table 4.1.

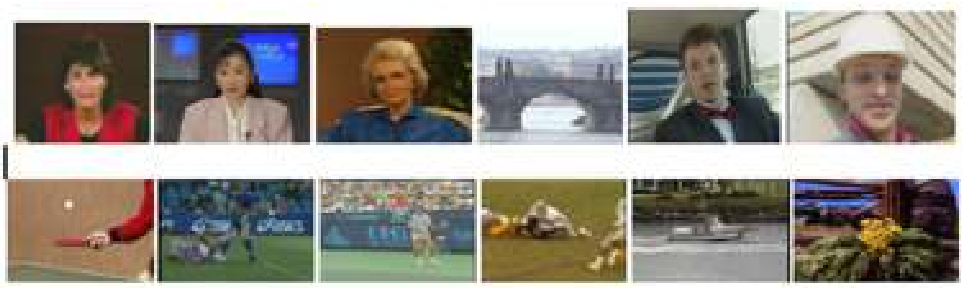


Figure 4.1. Snap shots of the video sequences

Table 4.1 Video sequences and their description

Sequence	Characteristics
Suzie	Close-up of woman talking on the phone, some head movement
Akiyo	Head & shoulders newscaster
Grandma	Grandma in front of the camera
Bridge-close	Charles Bridge (Karlv most) oldest bridge in Prague
Carphone	Man talking at the phone in a moving car, end with scene change
Foreman	Facial close-up followed by wide shot of construction site
Table tennis	Two players playing table tennis – lots of motion
Rugby	Outdoor rugby match: movement and colour
Stefan	Two players playing tennis – high motion
American football	American football scene, high motion
Coastguard	Still camera of a moving coastguard boat
Tempete	Moving Camera with flowers

4.4 Content dynamics

In this section the two dimensional content classification is presented from literature followed by the general impact of codec related parameters of SBR and FR on content dynamics and quality prediction.

4.4.1 Two dimensional content classification

The spatio-temporal complexity of the video varies from sequence to sequence. The coding of the video and the errors that occur when the video is transmitted over the networks is also very much dependent on the spatio-temporal complexity of the video sequence as shown in Chapter 3.It is desirable to classify videos according to their spatio-temporal complexity. Work presented [94] classifies videos according to their spatio-temporal complexity and proposes a spatio-temporal plane. In this plane each video signal (subject to short duration and homogeneous content) is presented as Cartesian point in the spatiotemporal plane, where the horizontal axis refers to the spatial component of its content dynamics and the vertical axis refers to the temporal one. The respective plane is depicted on Fig. 4.2.

Therefore, according to this approach, each video clip can be classified to four categories depending on its content dynamics, namely:

- Low Spatial Activity – Low Temporal Activity (upper left)
- High Spatial Activity – Low Temporal Activity (upper right)
- Low Spatial Activity – High Temporal Activity (lower left)
- High Spatial Activity – High Temporal Activity (lower right)

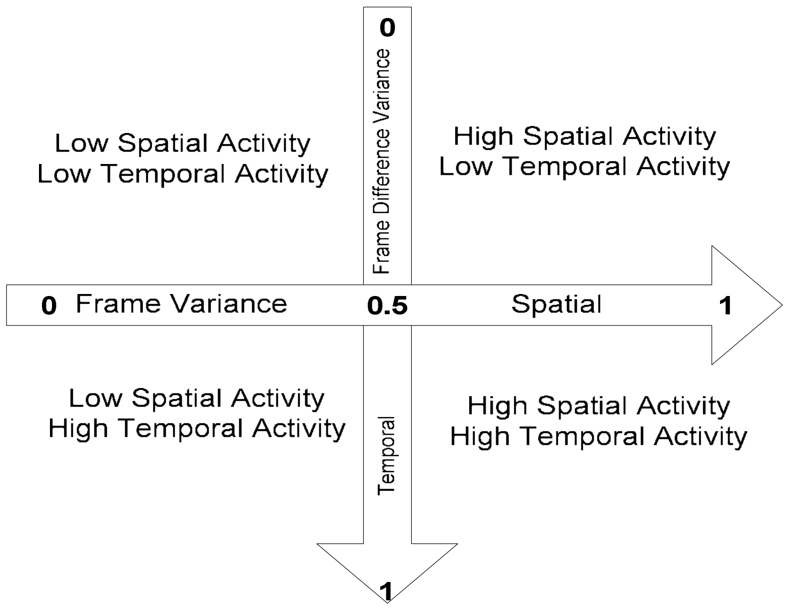


Figure 4.2. The Spatiotemporal grid

This plane is more representative of short video clips which are used in this thesis. For longer clips, the classification is less representative.

Spatial perceptual Information (SI) and Temporal Perceptual Information (TI) based on Sobel filter from ITU-T-Rec P.910 [1] was used in order to measure the complexity of the scenes given in Eqs. (4.1) and (4.2) respectively..

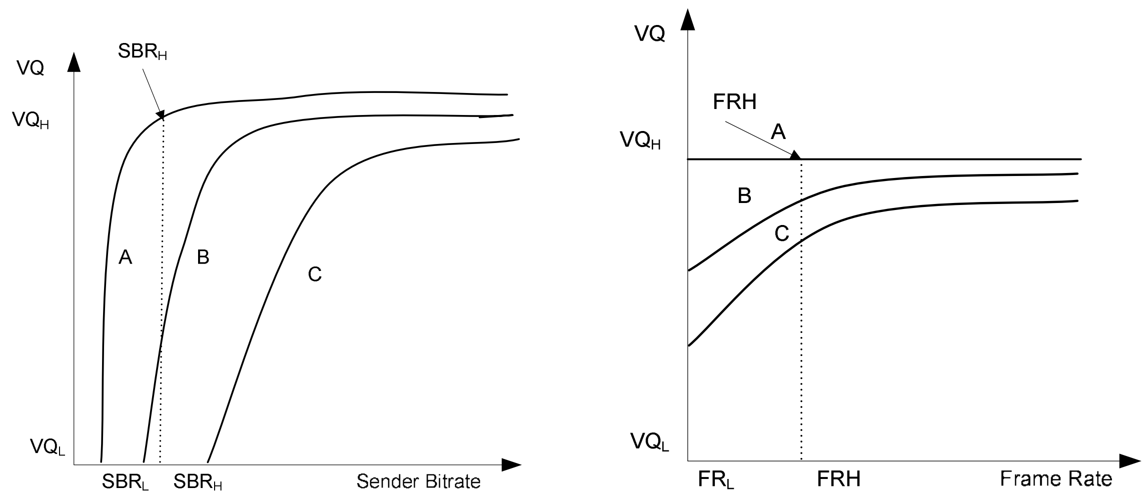
$$SI = \max_{time}\{std_{space} Sobel(F_n)\} \tag{4.1}$$

$$TI = \max_{time}\{std_{space}[\Delta F_n]\} \tag{4.2}$$

4.4.2 Impact of content dynamics on SBR and FR

In this sub-section, the impact of the spatiotemporal content dynamics on SBR and is presented. Figs 4.3 a and b show the impact of SBR and FR on video quality respectively.

From Fig. 4.3a, it can be seen that video quality only increases up to a point as the SBR is increased. Beyond that point, increasing the SBR only takes bandwidth but has no effect on quality. Curve ‘A’ represents video clip of low spatiotemporal dynamics, curve ‘B’ medium and curve ‘C’ high. It can be seen that as the spatiotemporal activity increases a higher SBR is needed to achieve the desired video quality, whereas, for low ST activity this SBR value is quite low. This is quantified with the data and the results are presented in Chapter 5, sub-section 5.6.1. Similarly, Fig. 4.3b shows the impact of FR on quality. It can be seen that FR has virtually no effect on low ST activity content, however, as SBR with higher ST activity, FR begins to have an impact.



(a) Video quality vs. sender bitrate (b) video quality vs. frame rate

Figure 4.3. Impact of content dynamics on sender bitrates and frame rates curves

It can be observed from Fig. 4.3a that in low sender bitrates curve A reaches a higher perceptual level compared to curve B depicting a sequence with higher spatiotemporal content. On the other hand, the curve C requires higher sender bitrate in order to reach a satisfactory VQ level. Nevertheless, curve(C) reaches its maximum VQ value more smoothly than in the low activity case. Full details of these curves can be found in [95].

Following the general pattern in Figs. 4.3a and b, it can be observed that the impact of the spatiotemporal activity on the sender bitrate and frame rate pattern is depicted very clear. It also shows two more important outcomes:

i) For video signals with low spatiotemporal activity, a saturation point appears, above which the perceptual enhancement is negligible even for very high encoding bitrates. However, frame rates do not have an impact on quality for the same videos.

ii) As the spatiotemporal activity of the content becomes higher, the respective perceptual saturation point (i.e. the highest VQ level) becomes lower, which practically means that video of high dynamics never reach a very high quality level. The low frame rates reduce the perceptual quality for the same videos.

4.5 Content classification based on ST feature extraction

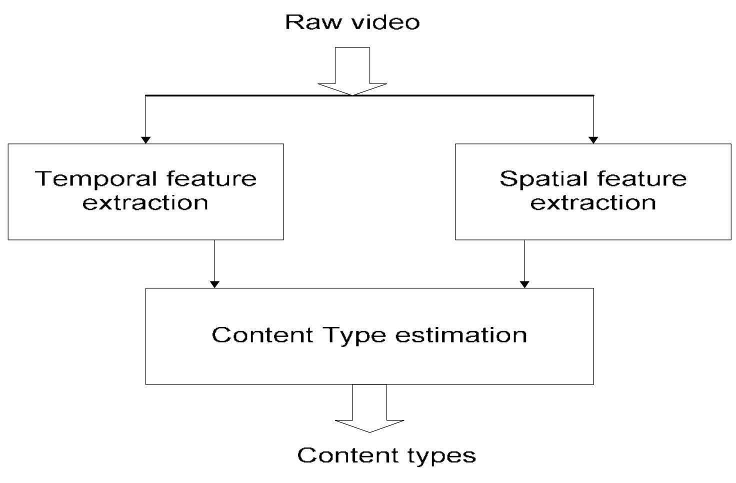


Figure 4.4. Content classification design

The chosen video sequences ranged from very little movement, i.e. on a fixed background, small moving area to fast moving sports clips. The choice of video sequences was to reflect the varying spatio-temporal activity of the content representative of typical content offered by content providers for video streaming applications e.g. news type of content or fast moving sports content. The 12 video sequences chosen are described in Section 4.3. The content classification was done based on the temporal and spatial feature extraction using well known tool called cluster analysis [91]. The block diagram of the content classification method is given in Fig. 4.4.

4.5.1 Temporal feature extraction

The temporal feature is extracted by extracting the movement part of the sequence... Backward and bidirectional prediction as specified by the ISO/IEC MPEG coders such as MPEG-4 part 10 is employed by Hybrid video compression standards. as report in [96] and [97]. The motion is estimated from the block of pixels called the macroblocks. The macroblock of the current video clip is subtracted from the reference video clip by the motion compensation prediction as pointed out by the appropriate motion vector. The movement in a video clip can be captured by the SAD value (Sum of Absolute Difference). In this thesis, the SAD values are used as temporal features and are computed as the pixel wise sum of the absolute differences between the two frames being compared and is given by Eq. (4.3).

$$\mathrm{SAD}_{n,m}=\sum_{i=1}^{N}\sum_{j=1}^{M}|B_n(i,j)-B_{\mathrm{m}}(i,j)| \qquad (4.3)$$

where B_n and B_m are the two frames of size N X M, and i and j denote pixel coordinates.

4.5.2 Spatial feature extraction

The spatial features extracted were the blockiness, blurriness and the brightness between current and previous frames [98].

Blockiness measures the blocking effect in video sequence. For example, in contrast areas of the frame blocking is not apparent, but they are clearly apparent in the smooth. The blockiness measure is calculated the visibility of a block edge determined by the contrast between the local gradient and the average gradient of the adjacent pixels [99] and is given by Eq. (4.4).

$$Blockiness = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\{\frac{1}{\mathrm{N}^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[x_m^n(i,j) - \bar{x}_m^n]^2\} \quad (4.4)$$

where $x_m^n(i,j)$ denotes the pixel value in location (i,j) of the mth block in the nth frame, $\bar{x}_m^n$ denotes the mean of the pixel values of the mth block in the nth frame, M denotes the number of blocks per frame, and N denotes the number of frames under investigation from the video sequence.

Blurring measurement is based on the measure of local edge expansions. The Sobel filter first calculates the vertical binary edge map . Following that, the local extrema (xp) in the horizontal neighbourhood of each edge point are identified, and the distance between these extrema (xp) is calculated. Blurring is computed as the average of the edge expansions for all edge points and is given be Eq. (4.5).

$$Bluriness = \frac{1}{N_e}\sum_{m=1}^{M}\sum_{n=1}^{N}|xp_1 - xp_2| \quad (4.5)$$

where N_e is the number of edge points. xp_1 and xp_2 are the local extrema in the horizontal neighborhood of each edge point.

Brightness (Br) is calculated as the modulus of difference between average brightness values of previous and current frames and is given by Eq. (4.6).

$$Br_{av\{n\}} = \sum_{i=1}^{N}\sum_{j=1}^{M}\left|Br_{av(n)}(i,j) - Br_{\mathrm{av(n-1)}}(i,j)\right| \quad (4.6)$$

Where $Br_{av(n)}$ is the average brightness of *n*-th frame of size N X M, I and j denote pixel coordinates.

4.5.3 Hierarchical Cluster analysis

12 video sequences were chosen reflecting very low spatial and temporal to very high spatial and temporal activity. Based on the table of mutual Euclidean norm in the joint temporal and spatial sense between pair of sequences, the dendrogram was created on the basis of a nearest distance in a 4-dimensional Euclid-space. The Euclidean distance is defined as the distance between the features of the 12 video sequences chosen. The dendrogram or tree diagram

constructed in this way classifies the content. The features (i.e. SAD, blockiness, blurriness and brightness measurements) extracted are given in normalized form. Fig. 4.5 shows the obtained dendrogram (tree diagram) where the video sequences are put in a group based on their mutual distances (nearest Euclid distance). The 4 features extracted from 12 video sequences are then input to the video content classification.

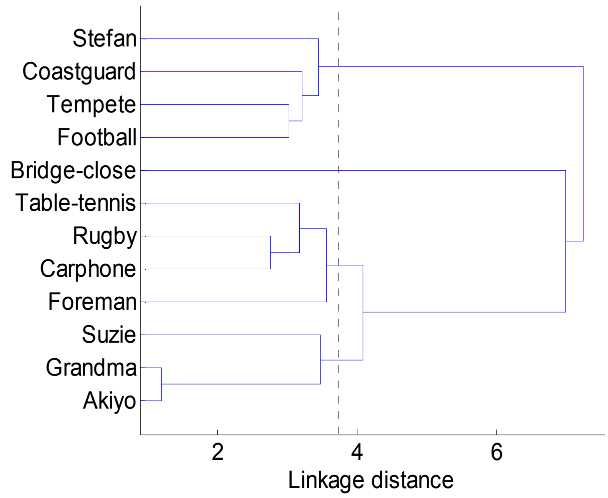


Figure 4.5. Tree diagram based on cluster analysis

According to Sturge's rule (k = 1 + 3.3logN), which for our data will be 5 groups. However because of the problems identified with this rule [100] the data (test sequences) was split at 38% from the maximum Euclid distance into three groups. (see the dotted line on Fig. 4.5) as the data contains a clear 'structure' in terms of clusters that are similar to each other at that point. Group 1 (sequences Grandma, Suzie and Akiyo) are classified as 'Slight Movement', Group 2 (sequences Carphone, Foreman, Table-tennis and Rugby) are classified as 'Gentle Walking' and Group3 (sequences Stefan and Football) are classified as 'Rapid Movement'. It was found that the 'news' type of video clips were clustered in one group, however, the sports clips were put in two different categories i.e. clips of 'stefan' and 'football' were clustered together, whereas, 'rugby' and table-tennis' were clustered along with 'foreman' and 'carphone' which are both wide angle clips in which both the content and background are moving. Also 'bridge-close' can be classified on its own creating four groups instead of three. But as it is closely linked with the first group of SM it was decided to put it in SM. This classification is further checked with k-means cluster analysis.

The cophenetic correlation coefficient, c, is used to measure the distortion of classification of data given by cluster analysis. It indicates how readily the data fits into the structure suggested by the classification. The value of c for the classification was 79.6% indicating a good classification result. The magnitude of c should be very close to 100% for a high-quality solution.

4.5.4 k-means Cluster Analysis

To further verify the content classification from the tree diagram obtained (Fig. 4.5) k-means cluster analysis was carried out in which the data (video clips) is partitioned into *k* mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation [83]. K-means computes cluster centroids differently for each measured distance, to minimize the sum with respect to the specified measure. k was specified to be three to define three distinct clusters. In Fig. 4.6 K-means cluster analysis is used to partition the data for the twelve content types. The result set of three clusters are as compact and well-separated as possible giving very different means for each cluster. Cluster 3 in Fig. 4.6 is very compact for the four video clips, whereas cluster 2 is reasonable compact. However, cluster 1 can be further divided into more groups. For example the video clip of bridge-close can be in a separate group. All results were obtained using MATLAB™ 2008 functions (Cluster).

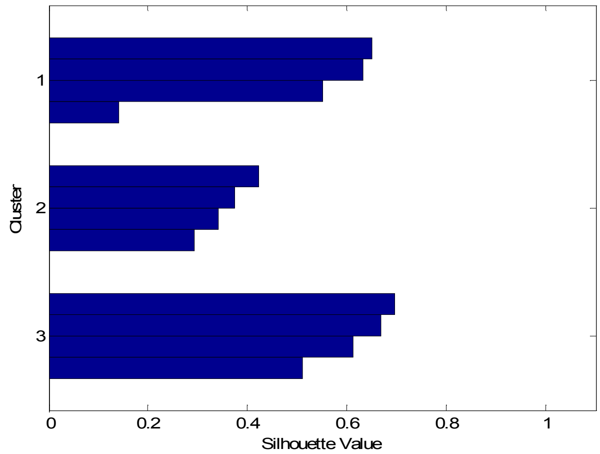
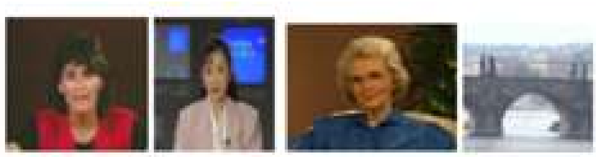


Figure 4.6. k-means of all contents types

4.5.5 Content type estimation

The three content types are defined for the most common contents for mobile video streaming are shown in Fig. 4.7a, b and c and described as follows:



(a) Snapshots of typical 'SM' content

(b) Snapshots of typical 'GW' content

(c) Snapshots of typical 'RM' content

Figure 4.7. Snapshots of all video contents classified using ST-feature extraction

Content type 1 – Slight Movement (SM): includes sequences with a head and shoulder type of movement (face) on a fixed background. See Fig. 4.7a.

Content type 2 – Gentle Walking (GW): includes sequences with a contiguous scene change at the end. They are typical of a video call scenario. See Fig. 4.7b.

Content type 3 – Rapid Movement (RM): includes a professional wide angled sequence where the entire picture is moving uniformly e.g sports type. See Fig. 4.7c.

4.6 Alternate method for content classification

4.6.1 Simulation set-up and data collection

The simulation test set-up is described in Chapter 3, section 3.3.3 and 3.3.4. However, instead of three types of video clips twelve video clips were used to generate the dataset. The video sequences ranged from slow to fast moving sports type of videos and are given in Section 4.3. A combination of FR, SBR, LBW and PER values were used to generate a total of 135 combination clips for each content type as shown in Table 4.2. In total there were 1620 test sequences.

Table 4.2 Dataset combinations over WLAN

<table>
<tr><td>Video sequences</td><td>Frame Rate fps</td><td>SBR (kb/s)</td><td>LBW</td><td>PER</td></tr>
<tr><td>Akiyo</td><td rowspan="12">10, 15, 30</td><td rowspan="8">18, 44, 80</td><td rowspan="12">32, 64, 128, 256, 384, 512, 768, 1000, 2000</td><td rowspan="12">0.01, 0.05, 0.1, 0.15, 0.2</td></tr>
<tr><td>Suzie</td></tr>
<tr><td>Bridge close</td></tr>
<tr><td>Grandma</td></tr>
<tr><td>Carphone</td></tr>
<tr><td>Coastguard</td></tr>
<tr><td>Tempete</td></tr>
<tr><td>Foreman</td></tr>
<tr><td>Rugby</td><td rowspan="4">80, 128, 512</td></tr>
<tr><td>Stefan</td></tr>
<tr><td>Table tennis</td></tr>
<tr><td>Football</td></tr>
</table>

Video contents are classified based on the Mean Opinion Score (MOS) obtained from parameters of SBR and FR in the application and PER in the access network level. A block diagram of the content classification model is given in Fig. 4.8. A well known multivariate statistical analysis called cluster analysis is used to classify the contents. Cluster analysis is chosen as it groups samples (video clips in this case) that have various characteristics into similar groups. Video MOS scores for all twelve video sequences obtained from objective video quality evaluation from the quality parameters of SBR, FR and PER are used as input to the statistical tool (cluster analysis) that classifies the video content into groups.

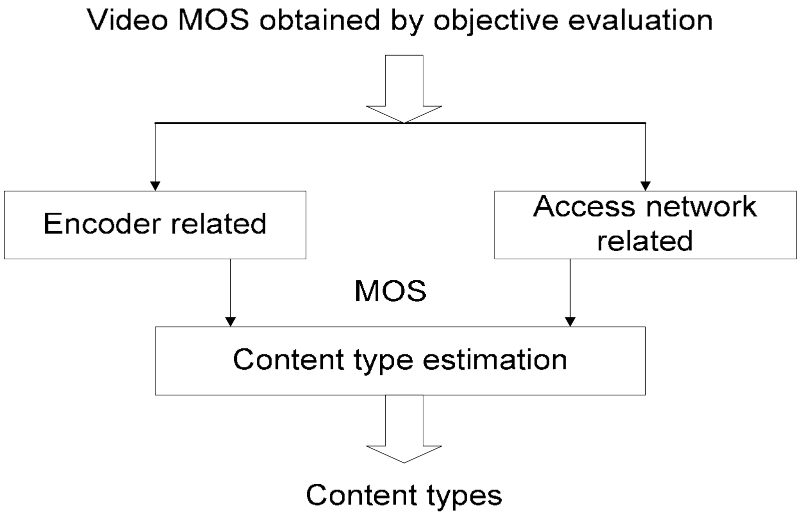


Figure 4.8. Content classification (alternate method)

4.6.2 Hierarchical cluster analysis

For the data, hierarchical cluster analysis was used in which samples that with the closest Euclid distance are put together. Fig. 4.9 shows the obtained dendrogram (tree diagram) where the video sequences are grouped together on the basis of their mutual distances (nearest Euclid distance).

According to Sturge's rule (k = 1 + 3.3logN), which for our data will be 5 groups. However because of the problems identified with this rule the data (test sequences) was split at 62% from the maximum Euclid distance into three groups. (see the dotted line on Fig. 4.9) as the data contains a clear 'structure' in terms of clusters that are similar to each other at that point. From Fig. 4.8 the video contents are divided into three groups of content types of Slight Movement (SM), Gentle Walking (GW) and Rapid Movement (RM). The spearman correlation coefficient is 73.29%. The correlation coefficient should be very close to 100% for a high-quality solution.

The correlation is not as good as in the previous method where contents are classified based on ST-feature extraction.

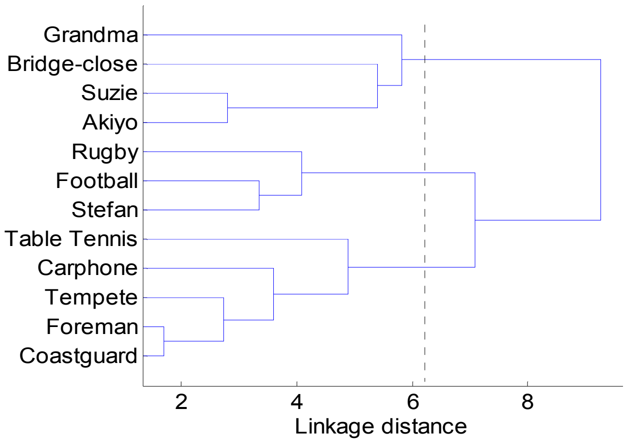


Figure 4.9. Tree diagram based on cluster analysis

4.6.3 k-means cluster analysis

To further verify the content classification from the tree diagram obtained (Fig. 4.9) k-means cluster analysis was carried out in which the data (video clips) is partitioned into *k* mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. K-means computes cluster centroids differently for each measured distance, to minimize the sum with respect to the specified measure [83].

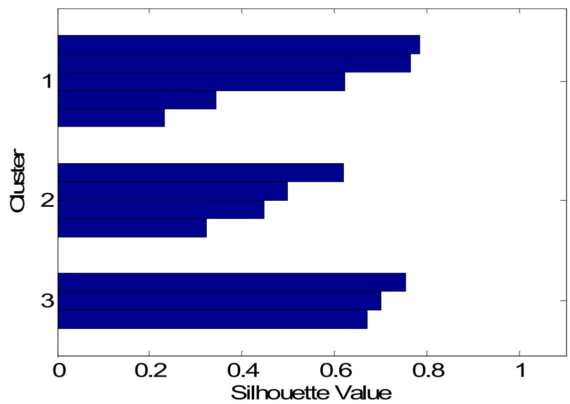
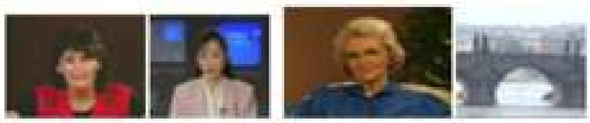


Figure 4.10. K-means cluster analysis

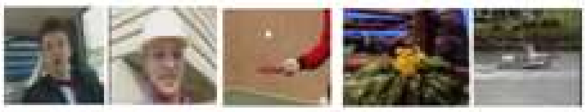
Again, k was specified to be three to define three distinct clusters. In Fig. 4.10 K-means cluster analysis is used to separate the data for the twelve video sequences. The result set of three clusters are as close together and well-detached as possible giving very different means for each cluster. Cluster 1 in Fig. 4.10 is very compact for three video clips instead of five. Clips of Table tennis and Carphone are slightly out of the cluster. They can be within their own cluster. Cluster 2 is reasonable compact and 3 is very compact. All results were obtained using MATLAB™ 2008 functions.

4.6.4 Content type estimation

Based on both hierarchical and k-means cluster analysis the content types are divided into three groups as shown in Fig. 4.11 and described below:



(a) Snapshots of typical 'SM' content



(b) Snapshots of typical 'GW' content

(c) Snapshots of typical 'RM' content

Figure 4.11 Snapshots of content using content classification method 2

Content type 1 – Slight Movement (SM): includes sequences with a head and shoulder type of movement (face) on a fixed background. See Fig. 4.11a.

Content type 2 – Gentle Walking (GW): includes sequences with a contiguous scene change at the end. They are typical of a video call scenario. See Fig. 4.11b.

Content type 3 – Rapid Movement (RM): includes a professional wide angled sequence where the entire picture is moving uniformly e.g sports type. See Fig. 4.11c.

It was found that the 'news' type of video clips were clustered in one group, however, the sports clips were put in two different categories i.e. clips of 'stefan', 'rugby' and 'football' were clustered together, whereas, 'table-tennis' was clustered along with 'foreman' and 'carphone' which are both wide angle clips in which both the content and background are moving. (see next section).

4.6.5 Comparison with ST grid

In this section the results obtained are compared to the ST grid shown in Fig. 4.2. Figure 4.12 shows the principal co-ordinates analysis also known as multidimensional scaling of the twelve content types. The function cmdscale in MATLAB™ is used to perform the principal co-ordinates analysis. cmdscale takes as an input a matrix of inter-point distances and creates a configuration of points [83]. Ideally, those points are in two or three dimensions, and the Euclidean distances between them reproduce the original distance matrix. Thus, a scatter plot of the points created by cmdscale provides a visual representation of the original distances and produces representation of data in a small number of dimensions.

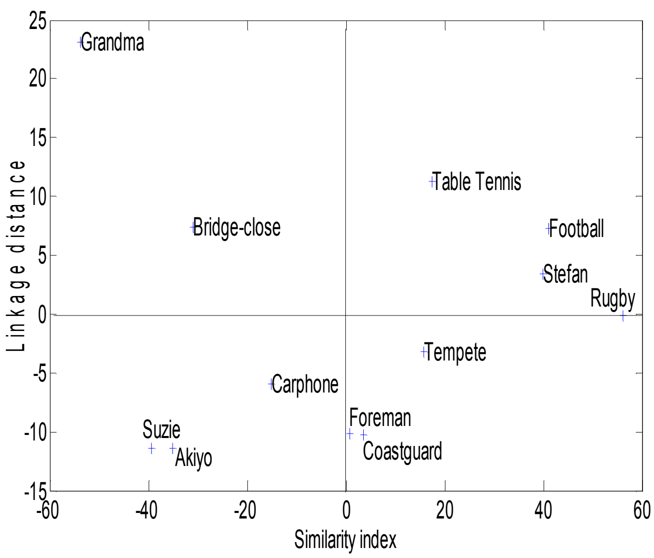


Figure 4.12. Principal co-ordinates analysis

In Fig. 4.12 the distance between each video sequence indicates the characteristics of the content, e.g. the closer they are the more similar they are in attributes. Comparing Fig. 4.12 to Fig. 4.2 it can be seen that classifying contents from the MOS scores (through objective video quality evaluation in our case), did group contents with similar attributes together e.g. contents of Football, Stefan Table-tennis and Rugby are high spatial and high temporal. However, according to the ST grid they should be in the bottom right hand side as opposed to the top right hand side, however rotating the grid by 270° shows that it can be fitted in the high temporal and high spatial feature contents into the top right hand side of Fig. 4.12. Fig. 4.12 will then be a better fit to the ST grid. The cophenetic coefficient was also much lower (~ 73%) indicating that the traditional method was a better fit. Also, the accuracy of classifying the contents this way can be difficult to quantify, as if the experiments are repeated, will it yield the same results, though it can be clearly seen that a pattern is formed as the PER increases and hence contents with similar ST features behave similarly to packet losses. More work is required to explore the accuracy of this method and has been left as a future work leading from this PhD thesis.

In the rest of the thesis content classification is carried out at the receiver side using the traditional method of ST feature extraction.

4.7 Summary

In this chapter two methods of content classification have been described. The first method classified the contents based on feature extraction, whereas, in the second method MOS values were used. It was concluded that the traditional method of ST feature extraction is more accurate. Although the second method gave interesting results as it grouped videos of similar attributes together. However, it may be difficult to reproduce that under different set of conditions. It is left as an area of future research leading from this thesis. Therefore, the rest of the thesis is based on the first method (ST-feature extraction method) for content type estimation carried out at the receiver side. Content type is an input to our models presented in Chapters 5 and 6.

Chapter 5

Regression-based Models for Non-intrusive Video Quality Prediction over WLAN and UMTS

5.1 Introduction

Multimedia content and services are growing exponentially across wireless access networks – both WLAN and UMTS. Digital videos are everywhere – from various hand held devices to personal computers. However, due to the bandwidth constraints of such networks, Quality of Service (QoS) still remains of concern. QoS is affected by parameters related to both the encoder (e.g. sender bitrate, frame rate, etc) and access network (e.g. block loss, jitter, etc) as was discussed in Chapter 3. The impact of these distortions is very much content dependent. For video applications to be successful over such access networks QoS is likely to be the major determining factor. In order to meet user's QoS requirements, there is a need to predict, monitor and if necessary control video quality. Non-intrusive models [101] provide an effective and practical way to achieve this.

Chapter 2 discussed the different objective (both intrusive and non-intrusive) and subjective video quality measurement methods. Intrusive methods are accurate and efficient however, they are impractical in real time monitoring. Hence, non-intrusive methods are preferred to intrusive analysis as they are more suitable for on-line quality prediction/control.

In this Chapter new regression-based models are developed for predicting video quality non-intrusively over wireless access networks of WLAN and UMTS. The prediction is from a combination of parameters associated with the encoder and access network for different types of content. The models are predicted in terms of the MOS obtained objectively (from simulation) and further from subjective tests.

This Chapter is organized is as follows. Related work on video quality modelling is introduced in Section 5.2. Section 5.3 outlines objective test set-up over WLAN and UMTS. Section 5.4 describes the subjective tests over UMTS. Section 5.5 presents the video quality prediction scheme. Section 5.6 presents the procedure for developing the models. The proposed models over WLAN are presented in Section 5.7. Section 5.8 presents the models over UMTS. Model comparison and validation with external databases is presented in Section 5.9. Section 5.10 summarizes the Chapter.

5.2 Related Work

The exponential growth of multimedia applications accessed via UMTS networks on mobile devices makes video quality prediction at the user level very desirable. Several studies on video quality prediction can be found in literature. Existing video quality prediction algorithms consider video content features or the effects of distortions caused by the encoder or network impairments. In addition, they are restricted over IP networks. However, with the growth of video services and applications over wireless access networks it is important to model losses that occur in the access network. Work presented in [16] have presented full reference video quality prediction models based on video content features for H.264 video. Reduced reference metrics presented in [102],[103] use raw video features to predict video quality. Whereas, works presented in [15],[104, 105]-[107] are from raw video features too, but the models are reference free. In [108] video quality for mobile applications has been evaluated. They chose content, codecs, bitrates and bit error patterns and found their impact on quality. They then present a no reference metric in [109] based on spatio-temporal features to estimate blur for images and video. Work presented in [110] compares three methods where spatio-temporal information and the impact of packet loss from the content is used to monitor video quality. In [111] a metric is presented that measures the temporal quality degradation caused by regular and irregular frame loss. In [11] video quality prediction models for H.264 video has been presented. The model presented is based on sender bitrate, frame rate and content types. In [112] perceptual metric for H.264 encoded panorama style video sequences has been presented. In [7] a theoretical framework is presented for MPEG4 video quality prediction. The framework considers sender bitrate in the application layer. In [8] a method to measure the quality of pictures of compressed videos based on an estimation of PSNR for H.264 encoded videos has been proposed. In [9] a model to measure temporal artefacts on perceived video quality in mobile video broadcasting services has been presented. They concluded that perceived quality for low spatio-temporal videos is affected adversely by frame rate decimation. In [10] a video quality metric based on quantization errors, frame rate and motion speed has been proposed. Metrics presented in [113] are to measure streaming video quality over windows media player. Work presented in [114] estimate the quality of H.264 encoded video sequences using a video decoder. They have used two parameters together (quantization parameter and contrast measure) within the H.264 decoder to give an estimation of subjective video quality. The prediction models presented in these works are from application layer parameters only (encoder based distortion and/raw content features).

In [12] network statistics e.g. packet loss, delay and jitter to predict video quality has been used, whereas in [13] a video quality measurement metric (rPSNR) developed from network packet loss conditions has been presented. Work presented in these papers considers network layer only. Work presented in [115] emphasize on emotional and physiological factors on video quality prediction and found direct relationships between video content features and physiological features. Similar to this in [116] psychological factors to estimate video quality of videophone services over IP networks for high sender bitrate videos has been proposed. In [117] a model called the V-factor using both transport and bitstream (raw video) information has been presented. They also present a review on objective video quality metrics and conclude that the work is still in its infancy. This makes our work on video quality prediction very timely. Work presented in [118] propose a video quality measurement metric based on $PSNRr_{,f}$-MOS_r. Their metric measures the performance loss due to damaged frames in a particular video sequence (f%) as well as to indicate the probability of a user experiencing a specified quality over the network (r%). They then find a linear relationship between PSNR and MOS.

5.3 Simulation set-up over WLAN and UMTS Networks

This section outlines the simulation set-up over WLAN and UMTS.

5.3.1 Data set generation and Experimental set-up over WLAN

The objective WLAN set-up was based on NS2 simulation. The combination of parameters along with the content types are given in Table 5.1.

Table 5.1 Dataset combinations over WLAN

Video sequences	FR (fps)	SBR(kbps)	LBW	PER (%)	MBL
Akiyo, Foreman	10, 15,30	18, 44, 80	256	1, 5, 10, 15, 20	1
Suzie, Carphone	10,15	50, 90			
Stefan	10,15,30	80,128,512	1000		
Football	10, 15	130, 384			

The sequences Akiyo, Foreman and Stefan are used for model training and Suzie, Carphone and Football are used for model validation over WLAN. The experimental set-up is given in Chapter 3, sub-sections 3.3.3 and 3.3.4. A total of 450 test sequences were generated for model training and 210 for model validation.

5.3.2 Data set generation and Experimental set-up over UMTS

The objective UMTS set-up was based on NS2 and OPNET simulation. The two simulation platforms were chosen to test whether the simulation environment had an impact on the test conditions gathered. The combination of parameters along with the content types are given in Table 5.2.

Table 5.2 Dataset combinations over UMTS

Video sequences	FR (fps)	SBR(kbps)	LBW	BLER (%)	MBL
Akiyo, Foreman	7.5,10, 15	48, 88,128	384	1, 5, 10, 15, 20	1, 1.75, 2.5
Suzie, Carphone	10,15	90,130			
Stefan	7.5,10,15	88,130,256			
Football	10, 15	130, 200			

The sequences Akiyo, Foreman and Stefan are used for model training and Suzie, Carphone and Football are used for model validation over UMTS. The experimental set-up is given in Chapter 3, sub-sections 3.5.3 and 3.5.5 over NS2 and 3.5.4 and 3.5.5 for OPNET. OPNET was used to analyze the specific impact of the UMTS error conditions into the perceived video quality, due to its accuracy of the implementation for the Radio-Link-Control (RLC) not-in-order delivery mechanism. Error simulated in the physical layer (BLER) is employed to generate losses at the link layer modelled with 2-state Markov model [54] with variable MBLs [82] to depict the various UMTS scenarios. A total of 567 test sequences were generated for model training and 224 sequences were generated for model validation over OPNET. NS2 was used due to its flexibility and based on the characteristics of the link bandwidth. A total of 450 test sequences were generated for model training and 210 for model validation over NS2. All the chosen test conditions were sent over the simulated network to generate test conditions with network impairments. This was specifically done as in literature only limited work on video quality assessment takes network errors into account. This enabled to observe the joint distortions caused by both application and access network impairments on end-to-end quality.

5.4 Subjective tests over UMTS Networks

This section describes the data collection and subjective tests set-up. The subjective tests were conducted using a PC at University of Plymouth (UOP). Additionally subjective tests were also conducted at University of Basque Country (EHU), Bilbao, Spain using mobile handset. The data collected from the handset-based tests have been used in this thesis to further verify the models and find the impact of devices.

5.4.1 Data collection

The dataset generation is exactly the same as sub-section 5.3.2. However, Table 5.2 is reduced to Table 5.3 as a subset of data is used for subjective tests. Frame rate was fixed at 10fps due to budget constraints of running the subjective tests. A total of 81 test sequences were generated for model training and 54 for model validation. The subjective tests were carried out via the Internet by the following URL:

http://www.tech.plym.ac.uk/spmc/staff/akhan/degraded_video.html

Table 5.3 Dataset combinations over UMTS for subjective tests

<table>
<tr><td>Video sequences</td><td>FR (fps)</td><td>SBR(kbps)</td><td>LBW</td><td>BLER (%)</td><td>MBL</td></tr>
<tr><td>Akiyo, Foreman, Stefan</td><td rowspan="2">10</td><td>48, 88,128</td><td rowspan="2">384</td><td rowspan="2">1, 5, 10, 15, 20</td><td rowspan="2">1, 1.75, 2.5</td></tr>
<tr><td>Suzie, Carphone, Football</td><td>90,130</td></tr>
</table>

The videos used in subjective tests were sent over OPNET [44] simulated UMTS network to create conditions with BLER of 1% to 20% as shown in Table 5.3. BLER of 20% related to

2% – 3% IP loss. Hence quality was not degraded beyond 20% BLER. The video test conditions are described in Table 5.3. The experimental set up is shown in Fig. 3.7. In total 135 test conditions were generated based on Table 5.3 from six videos. 81 were used for the training of the model and 54 were used for validation.

Subjective tests were carried out with the above test conditions in two countries over two terminals. In the University of Plymouth (UOP) Personal Computers (PC) were used in a lab, whereas in the University of Basque Country (EHU), Spain, mobile handset was used.

5.4.2 PC-based subjective set-up

The subjective quality assessment experiment pursues ITU-T Recommendations [40], and was carried out using the single-stimulus Absolute Category Rating (ACR) method. The ACR method has a five point quality scale as described in Chapter 2. Each degraded video is presented only once and is rated on an individual basis. The video sequences were presented to the viewers in a random fashion in such way that the video sequences were viewed in different presentation order by the subjects. The ratings of the video clips follow the ACR five point discrete scale from 'bad' (1) to 'excellent' (5). The MOS is then found according to [40] where all the ratings given by the viewers were averaged. Voting period for the subjects was not restricted by the time. Once the viewers chose their quality rating, they pressed the 'submit' button. This confirmed their choice. Before pressing the 'submit' button, the viewers were given the opportunity to change their mind before executing the final score. The distance to the monitor (viewing distance) was not fixed. All subjects were able to adjust their distance to the best that suited them.

The laboratory had calibrated 20-inch computer LCD monitor (Philips 200 WB7) to show the video sequences. The display had a built-in resolution of 1280 x 1024 pixels and the colour quality was selected as highest (32 bit). The room had a white background. The subjects rated the final score using the computer mouse by clicking on the 'submit' button.

The total number of subjects were 20, it included 11 males and 9 females. They were naïve subjects as they had not participated in a video quality assessment before.. This complies to the minimum number of viewers specified by ITU-T Recommendations [40]. The age range of 14 participants was between 18 to 25, 4 were between 25 to 30 and 2 were over 35. Subjects were recruited from within the University. The experiment started by presenting the subjects first with three training sequences that were different to the test sequences.. The experiments were divided into two sessions with a comfortable 10-15 minutes break between them over three days. This adhered to the ITU-T recommendation of time period not exceeding half an hour. An informal survey was conducted after the tests regarding the length of the study, fatigue during the tests, etc. which concluded that all participants did not experience any fatigue or uneasiness during the course of the tests. All the degraded video sequences were assessed by all the subjects that took part in the experiment.

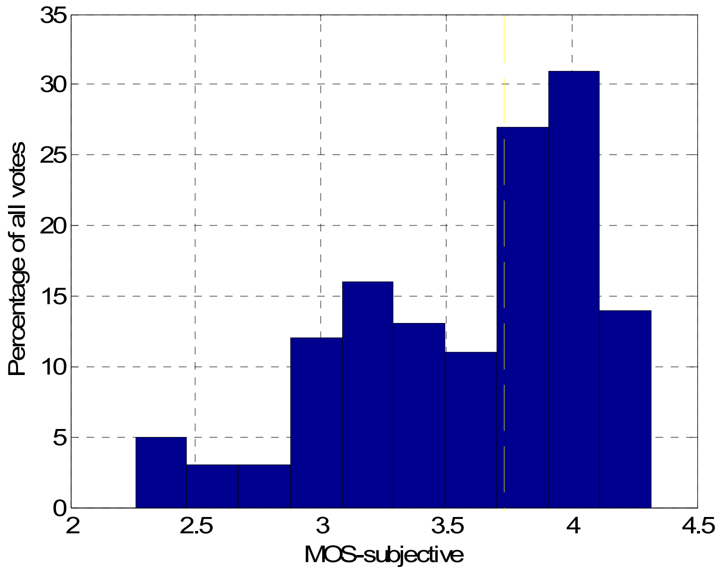


Figure 5.1. Histogram of subjective MOS (median 3.7)

The obtained MOS data was scanned for conflicting results. The ITU-T [40] criteria to reject 'bad results' which rejected three subjects. The scores from the rest of the subjects were averaged to compute the overall MOS for each test condition. Fig. 5.1 shows the histogram of subjective quality ratings. The median quality in the experiment is 3.7 and is represented by the dashed line in Fig. 5.1.

5.4.3 Handset-based subjective set-up

The handset-based tests were conducted in Spain. The experimental design for the subjective video tests is based on ITU recommendations [40],[119] . The video clipwere presented to 20 different people, that included 13 males and 7 females. All video sequences were displayed in a mobile handset according to [120] and users were asked to hold it on their hands. The device used in the tests is a Nokia N95-8Gb, which provides a 320x240 screen resolution as shown in Fig. 5.2. The tests are carried out with CorePlayer™ Mobile software due to its flexibility for using playlists.

Figure 5.2. Video test on Nokia N95-8Gb

There was a training session in which the subjects were given some video clips which were different to the ones used in the experiment. After the training session, subjects were shown the video clips for the experiment. The experiment was carried out over three days and the video clips were randomized and checked for coherence [119]. Breaks were given in between the experiments. Fig. 5.3 shows the histogram of subjective quality ratings. The median quality is shown by the dashed line in Fig. 5.3.

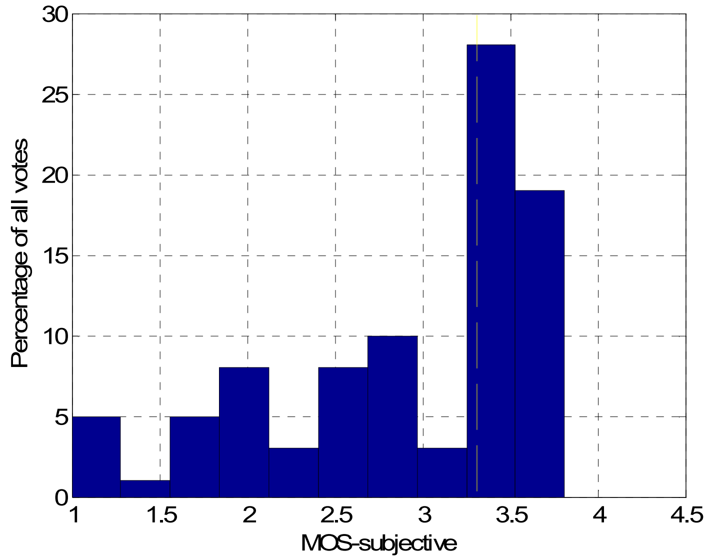


Figure 5.3. Histogram of Handset-based MOS dataset (median at 3.3)

The MOS results obtained for the test conditions defined were made publicly available to the research community at:

http://www.tech.plym.ac.uk/spmc/staff/akhan/mos_scores.html [41]

5.4.4 Comparison of MOS from two devices – PC and Handset

The histogram of the MOS values shown in Figs. 5.1 and 5.3 obtained from two devices show that subjects gave a relatively higher score for nearly all video clips that were viewed on PC as compared to handset. It was found that the range of subjective MOS were dependent on the end device. A MOS range of 2.3-4.4 from the PC was achieved, whereas the mobile handset gave a range of 1-3.8. This showed that with 1% BLER viewers rated the same video clip higher on a PC compared to mobile handset. Not surprisingly, the low motion content received the best ratings both on PC and handset. For fast moving sports type of clips the user acceptability was very low. However, it was interesting that the same degraded clip e.g Stefan at 20% BLER was scored at 2.3 over PC compared to 1 over handset for an SBR of 88kbps. Subjective feedback highlighted that users found the lack of rescale on mobile device quite annoying and hence rated the videos low compared to those who viewed them on PC. Also the distance to screen affected the viewers too. However, results from both countries and both devices followed the same pattern of MOS degradation with the increase on transmission errors. Figure 5.4 shows the correlation between MOS from handset and PC. There was a correlation of 83.58% which was quite low. For low SBRs the difference of quality ratings between PC and handset is less. This is because at low SBR quality is relatively poor. However at higher sender bitrates e.g. 128bkps there is a larger reduction in quality over handset as compared to PC due to the reduction in size. The other factor is the cultural differences between the two countries. That would have an impact on user expectation and scoring.

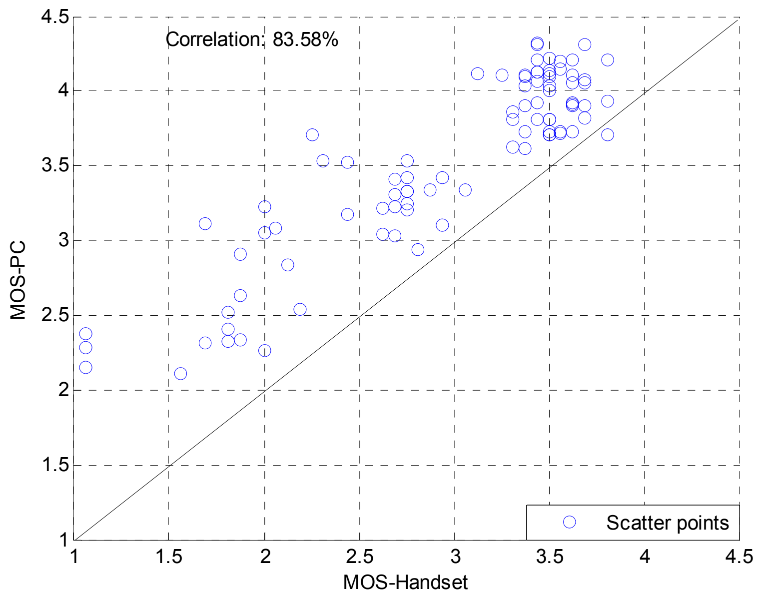


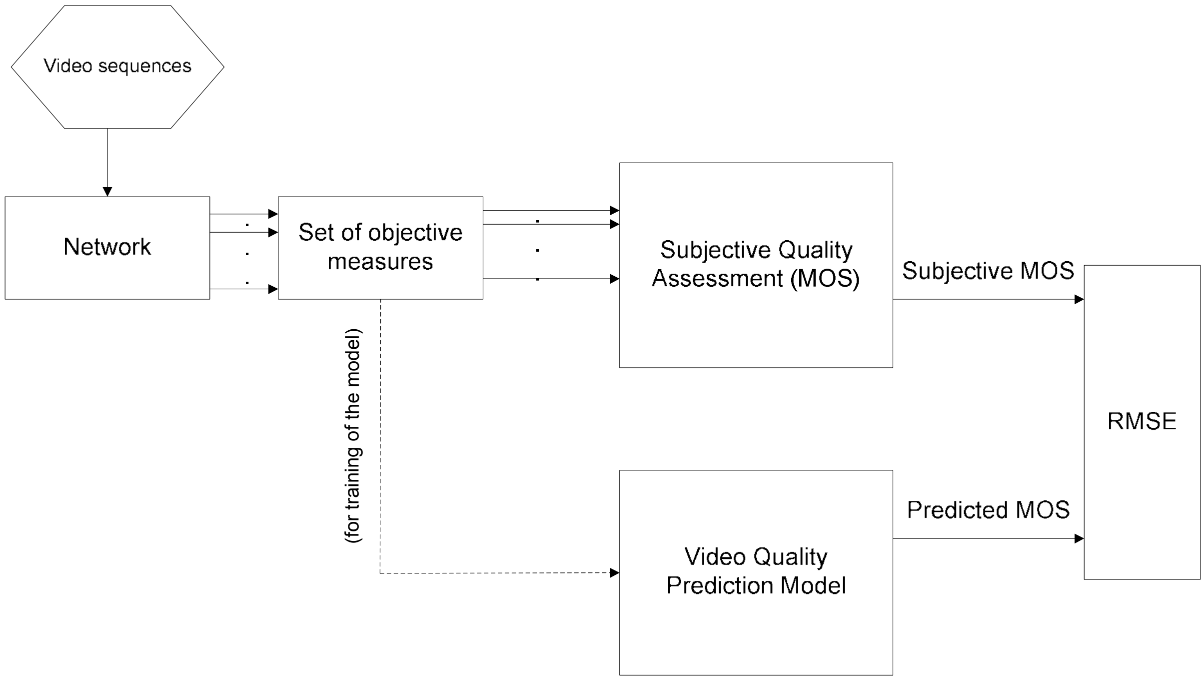
Figure 5.4 Correlation between MOS from handset and PC

5.5 Novel non-intrusive Video Quality Prediction Models

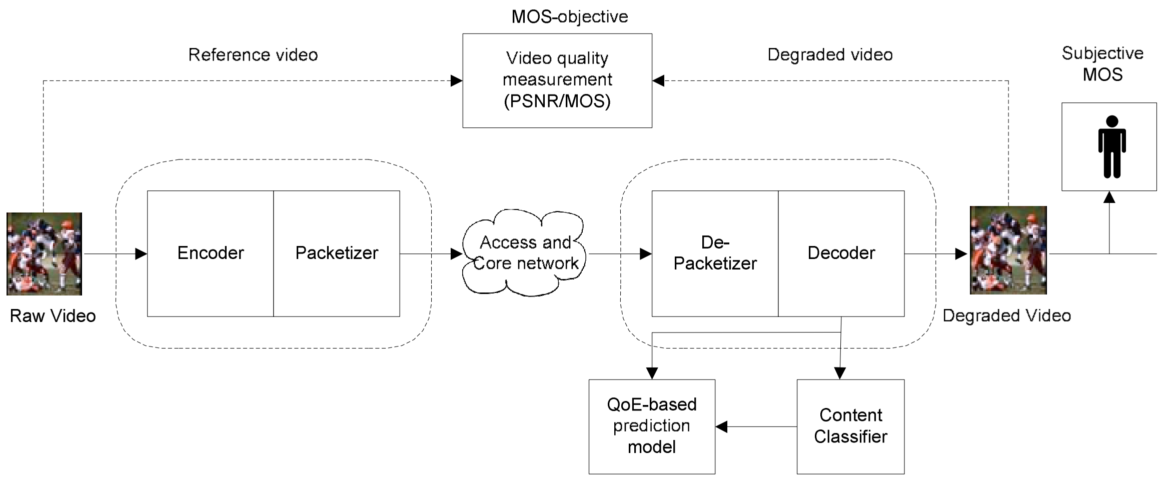
Figures 5.5 a & b depicts a simplified, conceptual diagram for developing the non-intrusive prediction model of video quality over WLAN/UMTS networks. Fig. 5.5a gives the block diagram of the video quality assessment model. Fig. 5.5b shows the end-to-end quality measurement. The video quality is predicted using either the non-linear regression models or ANN-based learning models in terms of the MOS, non-intrusively. The video quality is predicted from a combination of parameters associated from network parameters (e.g. packet/block loss and MBL), codec related parameters (e.g. SBR, FR) and content types. In real time, the parameters are extracted from the header of the decoder (e.g. RTP) to give the content related information, based on which video contents are classified.

. The regression models presented in this thesis uses (a) PSNR to MOS conversion from simulation and (b) subjective MOS results to provide an objective measure of video quality (Measured in MOS) which is then used to generate appropriate data for curve fitting (for regression models) as shown in Fig. 5.5a or for ANFIS based neural network training (see Chapter 6 for details). The benefits of this method for non-intrusive video quality prediction are:

- It is generic and based on end-to-end, intrusive measurement of video quality. Thus, it can be easily applied to other applications, such as audiovisual (and hence quality of multimedia). Both the ANFIS-based neural network models will have to be re-trained to account for the new data and the regression based models would have to be re-derived based on the new data.
- Subjective tests consume time and are expensive. Accurate objective models can avoid subjective tests.
- The models can be re-derived or re-trained with new QoS parameters.



(a) Block diagram of the video quality prediction model



(b) End-to-end video quality prediction

Figure 5.5. Conceptual diagram of non-intrusive video quality

The non-linear regression and neural network models presented in this thesis are generic and have applications in the optimization of content and network QoS control by adapting the video quality to changing access network conditions
These applications are explored in Chapter 7.

5.6 Procedure for developing non-intrusive video quality prediction models

This section outlines the procedure for developing the non-intrusive prediction models. First, the relationship between the encoder parameter of Sender Bitrate (SBR) with end-to-end quality is found and hence the minimum acceptable SBR for each type of content is found. Next, the acceptable Packet Error Rate (PER) over WLAN is found for each type of content. These relationships are found in terms of PSNR. Finally, the mathematical relationship between each QoS parameter and end-to-end quality (in terms of the MOS-subjective) is found.

5.6.1 Relationship of encoder parameter of SBR with quality (PSNR)

The following research question was investigated to find the relationship of the encoder related parameter of SBR with quality. Quality is measured in terms of PSNR at this stage. At a later stage PSNR to MOS conversion from Evalvid is carried out. The conversion table is given in Chapter 3, Table 3.2.

What is the minimum send bitrate for all content types to meet communication quality for acceptable QoS (PSNR >27 dB) as it translates to a MOS of greater than 3.5 [1]

ITU-T P.910 [1] defines MOS of 3 as fair and MOS of 4 is considered good. Therefore, ITU-T P.910 recommends a MOS of 3.5 as acceptable for video communication over IP/wireless networks.

To answer the research question, the frame rate was fixed at 30fps.

The sender bitrate versus PSNR curve is shown in Fig. 5.6 for all contents. From Fig. 5.6 it is observed that there is a minimum sender bitrate for acceptable quality (PSNR > 27dB) for all content types. For high definition IPTV applications PSNR of 32dB is recommended. Therefore, in Fig. 5.6 32dB has been chosen as minimum acceptable PSNR as compared to 27dB to illustrate the point of optimizing bandwidth. Also there is a maximum send bitrate for the three content types that gives maximum quality (PSNR > 38dB). For example for the content category of SM, sender bitrate of 30kbps or more gives a maximum PSNR of 38dB. However, in RM higher sender bitrates are required for maximum quality i.e. > 370kb/s. From Fig. 5.6 it can be derived that when the sender bitrate reduces below a certain threshold, which is depends on the video content, then the quality becomes 'bad'. Moreover, when the sender bitrates are above a specific threshold then quality improvement is not significant depending on the spatio-temporal activity of the video clip.

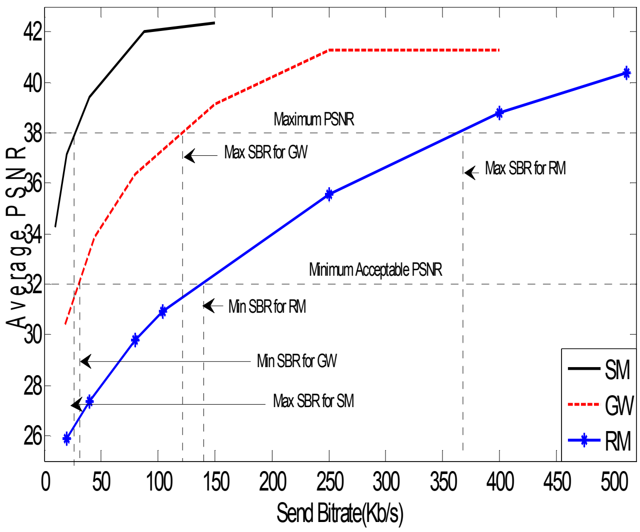


Figure 5.6. PSNR vs Send Bitrate for the three contents

The send bitrates ranged from 18kb/s to 384kb/s. One video clip was chosen from each category. A minimum sender bitrate was suggested for all three categories that achieve an average PSNR values of higher than 27dB for the video content types as it translates to a

MOS of greater than 3.5 [1] which is an acceptable score for the telecommunication industry. It was found that for slow moving contents (contents category of SM), 18kbps gave acceptable quality of 32dB. For content type of GW (medium movement), 32kbps gave acceptable quality of 32dB. For fast moving contents (content category of RM) it was found 140kbps as acceptable SBR as it gave a PSNR of 32dB. These quality measurements were taken with no packet losses. Finally, the relationship between PSNR and quality was found for the three content types. This is given by Equations 5.1 to 5.3. A logarithmic relationship was found between SBR and video quality (PSNR). The equations were derived in MATLAB by finding the best fit to the data represented in Fig. 5.6.

$$PSNR_{SM} = 27.64 + 3.08\ln(SBR) \tag{5.1}$$

(R^2=96%, RMSE=0.664)

$$PSNR_{GW} = 19.75 + 3.76\ln(SBR) \tag{5.2}$$

(R^2= 97%, RMSE =0.702)

$$PSNR_{RM} = 10.64 + 4.61\ln(SBR) \tag{5.3}$$

(R^2= 96%, RMSE=1.1)

5.6.2 Relationship of PER with quality over WLAN

The following research question was investigated to find the relationship of wireless access network parameter of PER with quality over WLAN:

What is the acceptable packet error rate for all content types for streaming MPEG4 video and hence, find the threshold in terms of upper, medium and lower quality boundary at which the users' perception of quality is acceptable (>27dB)?

To answer this research question, all videos were encoded at 256kbps, 30fps. Two metrics were chosen as PSNR and Q.

PSNR

Video quality is measured by taking the average PSNR over all the decoded frames across network PER from 0.01 to 0.2 (20%). All videos were encoded at a sender bitrate of 256kb/s. Twelve content types were chosen as shown in Fig. 4.1 (Chapter 4). Fig. 5.7 show the average PSNR vs the PER for all 12 video clips. It shows that the average PSNR is better for slight movement compared to gentle walking which in turn is better than rapid movement which shows the dependence on content type. From the results, it was found that for slight movement the video quality stays above the threshold of PSNR > 27dB (MOS >3.5) for upto 20% packet loss. However, for gentle walking and rapid movement that value drops to 10% and 6% respectively.

It is observed from Fig. 5.7 that the drop in video quality is much higher for fast moving contents compared to that of slow moving contents. E.g. for 'Akiyo' at 0.01 PER the PSNR is 44dB and at 0.2 (20%) PER it is 27.67dB. However, for 'Football' it is 33dB for a PER of 0.01 and 20dB for PER of 0.2. Even though the percentage drop in quality is more or less the same, 20dB is unacceptable for communication standards. This can be furthered explained by the fact that the bitrate was fixed at 256kb/s. If the bitrate is varied then the impact of packet error rate is much greater on fast moving contents.

The PSNR of SM was found to be 27.67dB for 20% packet loss, for GW this value was 28.102dB for 10% packet loss and for RM a range of 25.57db – 27dB was found for 6% packet loss. Therefore, 6% packet loss was taken to be acceptable. However, all of these three values as shown later (Fig. 5.10) give unacceptable quality.

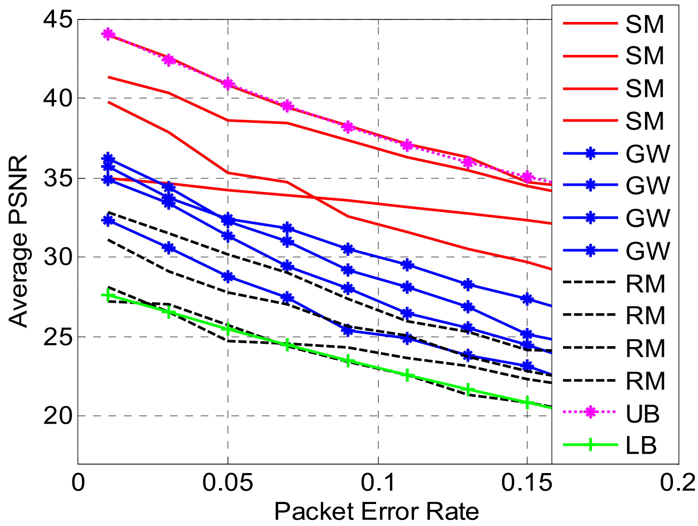


Figure 5.7. Packet Error Rate vs Average PSNR

Further, an upper, medium and lower boundary for PSNR as a function of PER for the three content types of SM, GW and RM was derived and hence know the threshold for acceptable quality in terms of the PSNR for the three content types with 95% confidence level and goodness of fit of 99.71% and Root Mean squared Error (RMSE) of 0.3235 is given by Equations (5.4), (5.5) and (5.6):

SM: $PSNR = 122.3(PER)^2 - 88.36(PER) + 42.6;$ $\quad PER \leq 20\%$ *(5.4)*

GW: $PSNR = 64.9(PER)^2 - 73.75(PER) + 34.43;$ $\quad PER \leq 10\%$ *(5.5)*

RM: $PSNR = 76.8(PER)^2 - 68.87(PER) + 31.43;$ $\quad PER \leq 6\%$ *(5.6)*

Q-value

The above research question was further addressed in terms of Q [4] instead of PSNR.

Fig. 5.8 shows the decodable frame rate (Q) of all 12 contents and shows that Q is higher when the PSNR is higher for all the video clips. In comparison to Fig 5.7 the decodable frame rate does not directly compare to the PSNR. However, from the results it was found for the average PSNR for 'slight movement' and it did not correspond to a higher value of Q. This is because the Q value is derived from the number of decodable frames over the total number of frames sent by a video source i.e. it is sensitive to the number of frames and packets lost.

Therefore, as the content becomes more complex the video quality was expected to degrade more for less I-frames lost compared to that of simpler contents. Hence, it is concluded that for slight movement 20%, for gentle walking 10% and for rapid movement 6% packet loss is acceptable.

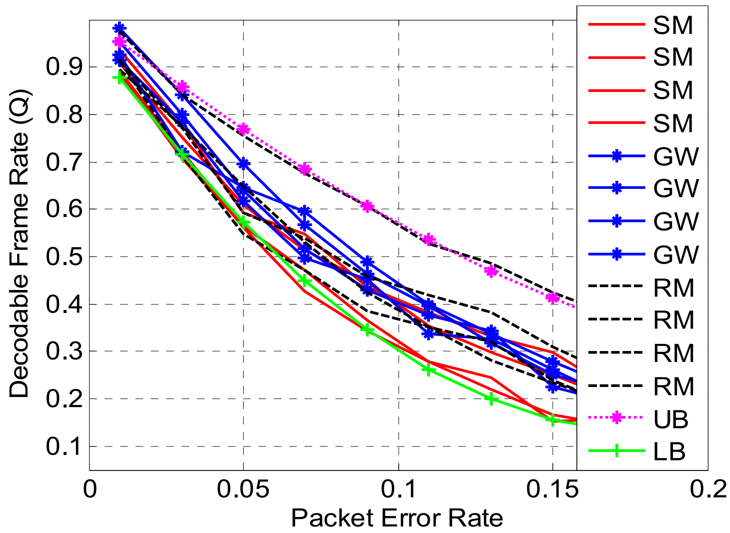


Figure 5.8. PER vs Q for all content types

Further, an upper, medium and lower boundary was derived for Q value as a function of PER for the three content types of SM, GW and RM and hence know the threshold for acceptable quality in terms of the Q value for the three content types with 95% confidence level and goodness of fit of 99.71% and RMSE of 0.0117 is given by the Equations (5.7), (5.8) and (5.9):

SM: $Q=19.89(PER)^2 - 8.03(PER) + 0.967$; $PER \leq 20\%$ *(5.7)*

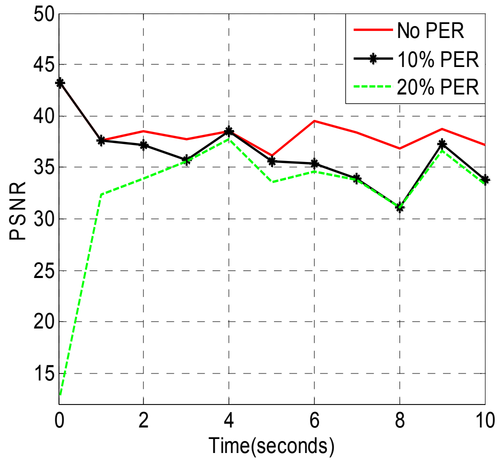
GW: $Q=18.09(PER)^2 - 7.88(PER) + 1.02$; $PER \leq 10\%$ *(5.8)*

RM: $Q=13.84(PER)^2 - 6.5(PER) + 0.975$; $PER \leq 6\%$ *(5.9)*

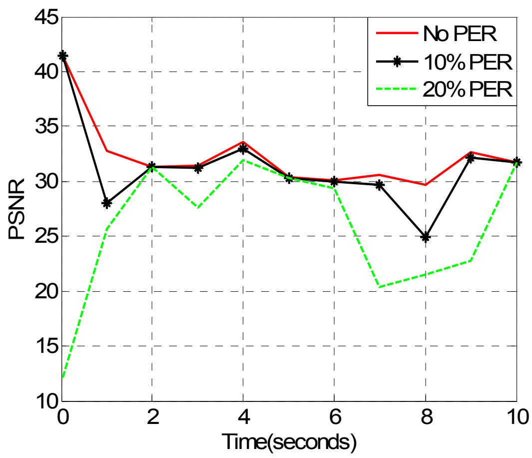
The difference in results between PSNR and Q as a metric were not very significant, all future results were recorded using PSNR only. Further, PSNR to MOS conversion was obtained as well.

The relationship between the PSNR over the entire duration of the sequences for all three content types was investigated.

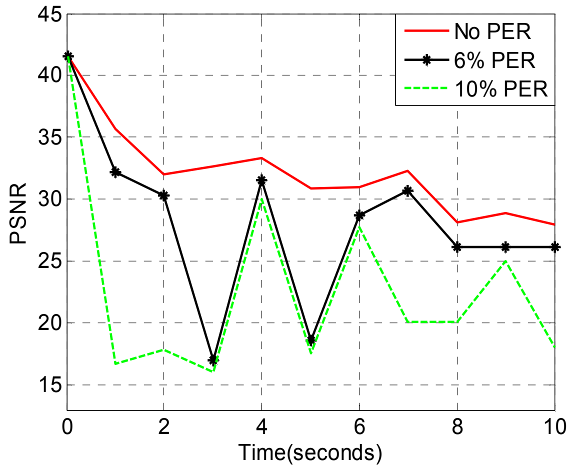
In Fig. 5.9a the source of effects caused by packet errors was investigated over the entire duration of the sequence. For ‘slight movement’ the PSNR values are compared for no packet loss to 10% and 20% packet loss. With higher losses, the error occurs in the B-frames and propagates to the P-frames as expected. Two effects are observed, the PSNR decreases over the entire duration and the second a more ragged response curve when packet errors of 10% and 20% are introduced. It is also observed that for a sender bitrate of 32kb/s the video quality is still acceptable for 20% packet loss.



(a) PER effects for SM for 32kb/s



(b) PER effects for GW for 80kb/s



(c) PER effects for RM for 256kb/s

Figure 5.9. PSNR vs Time

Fig. 5.9b shows the effects of no packet loss, 10% and 20% packet loss for ‘Gentle walking’ at a sender bitrate of 80kb/s. Again as previously mentioned the video quality reduces over the time duration and it is observed a much bigger loss in quality as the packet loss increases to 20%.

Whereas, from Fig. 5.9c in ‘rapid movement’ the video quality degrades fairly quickly with the increase of packet error rate i.e. for 10% packet loss the video quality is completely unacceptable.

While PSNR is not a good predictor of the visual quality, it can serve as a detector of clearly visible distortions. It can be observed, however that the perceived quality degradation increases in the duration of the sequence. Due to the auto-correlation of the time series (each sample is dependent on the previous and following sample) the values are not independent. It is also observed that as the scene activity in the video sequence becomes more complicated e.g. for ‘rapid movement’ at 20% packet loss the quality is completely unacceptable deteriorating at a much faster speed. All degraded video clips can be found in [121].

Fig. 5.10 shows that visually the quality of SM, GW and RM is unacceptable at 20%, 10% and 6% packet loss for some frames. Also from Table 5.4 it was observe that even though PSNR value is acceptable (MOS>3.5) for all three content types, however, the end-to-end perceptual quality is unacceptable. From Fig. 5.10a, the PSNR at 3.4 seconds for SM shows a value of 35dB, whereas the frames (101-103) from Fig. 5.10a show that the perceptual quality does not follow for those frames. Similarly, for GW at 5.2s (Fig. 5.10b) the PSNR is 30dB and for RM at 3.2s it is 17dB (Fig. 5.10c). The PSNR values of GW and RM reflect the perceptual quality better compared to SM. Further from Table 5.4 it can be seen that for SM, more B-frames are lost compared to GW and RM. B-frames affect the quality least in MPEG4 GOP. I-frames take priority, then P-frames and finally B-frames. Also the values of Q correlate well with PSNR for GW and RM. However, for SM it does not. Q-value for SM actually shows that at 20% the quality is less than acceptable compared to that of PSNR. Therefore, subjective tests were carried out (see Section 5.4) as PSNR is not a good indicator of perceptual quality.

(a) Frames 101-103, PER @ 20% for SM encoded at 32kb/s

(a) Frames 156-158, PER @ 10% for GW encoded at 80kb/s

(b) Frames 96-98 , PER @ 6% for RM encoded at 256kb/s

Figure 5.10. Perceptual quality comparison for the 3 content types at PER 20%, 10% and 6%

Table 5.4 summarizes the findings of Figs. 5.9 and 5.10 and outlines the PSNR and Q values for acceptable quality at 20%, 10% and 6% PER for all three content types in terms of the I, P and B frames lost. It is observed from Table 5.4 that for content type of SM the Q value is much lower compared to that of the PSNR. It shows that visually the quality is much lower at 20% packet loss rendering PSNR to be not a very good predictor of visual quality. For SM, Q-value out-performs the PSNR.

Table 5.4 PSNR and Q values for three content types

	Packets Lost (PER)	I-frames lost	P-frames lost	B-frames lost	PSNR	Q-value
SM	20%	8	14	43	27.67	0.458
GW	10%	8	7	22	28.103	0.602
RM	6%	8	11	12	25.57	0.615

5.6.3 Mathematical relationships of QoS parameters over UMTS

In this section the mathematical relationship followed from the analysis of Chapter 3 is described between the video quality (in terms of MOS) and individual QoS parameters related to the encoder, access network and content types. These relationships are used in Sections 5.7 and 5.8. In this section, subjective MOS has been used.

The regression-based video quality prediction model is developed for all content types from the relationships found for video applications over WLAN and UMTS networks. The experiment takes into account six test sequences, divided in two groups: akiyo, foreman and stefan are used for training the model, while carphone, suzie and football available are devoted to the validation of results. The video sequences represent content with low Spatio-Temporal (ST) to high ST features as classified in Chapter 4. As the transmission of video was for hand held devices and mobile handsets, all the video sequences were of QCIF resolution (176x144) and encoded in H.264 with Baseline Profile at 1.2 level for UMTS with an open source JM software [46] encoder/decoder. The considered frame structure is IPPP for all the sequences, since the extensive use of I frames could saturate the available data channel. The combination of parameters chosen are given in Section 5.4 (subjective testing). The functional block of the proposed model is shown in Fig. 5.11. The parameters associated with the encoder are Frame Rate (FR) and Sender Bitrate (SBR), Content Type (CT) and the parameters associated with the access network are access network (WLAN/UMTS) are PER//BLER modeled with 2-state Markov model with variable MBL.

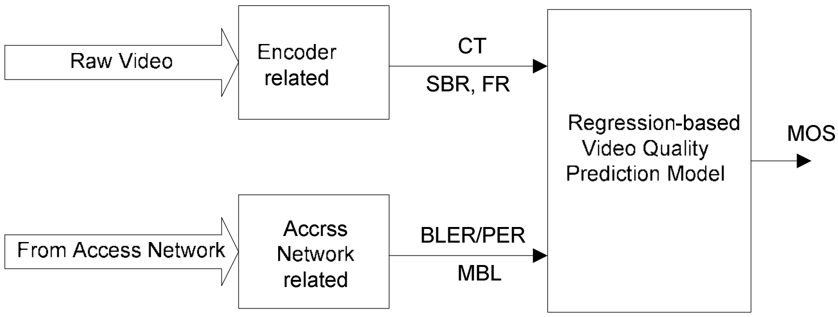


Figure 5.11. Functional block of proposed regression-based model

<u>Relationship of Content type and Encoder Parameters</u>

In addition to the 3D graphs plotted given in Chapter 3, 2D graphs were plotted to analyze the relationship of MOS with sender bitrate, content type, block error rate and mean burst length as given in Fig. 5.11. Based on the relationships of the QoS parameters the following function was established for estimating video quality. Therefore, the overall MOS is then given by Eq. (5.10).

$$MOS = \frac{k + f_A}{f_P} \tag{5.10}$$

Where k is a constant, f_A is measured in terms of SBR and CT and f_P is measured in terms of BLER and Mean Burst Length (MBL) for 2-state Markov model.

Mathematically $f_A = f_A(SBR, CT)$ and $f_P = f_P(BLER, MBL)$

CT is found by extracting temporal features of Sum of Absolute Differences (SAD) and spatial features of edge, blurriness and brightness, as described in Chapter 4 thus giving CT by Eq. (5.11).

$$CT = f(SAD, edge, blurriness, brightness) \qquad (5.11)$$

Based on the extracted spatio-temporal features, cluster analysis is carried out based on the Euclid distance of the data to determine the content type. Therefore, video clips in one cluster have similar content complexity. The content classification in this thesis is carried out offline. The block diagram of the content classification carried out at the receiver side is shown in Fig. 5.12. The learned classifier (cluster analysis) takes the extracted features for each new video as input then predicts its most likely type. Once the CT is predicted, then it is used as an input to the video quality prediction model (see later).

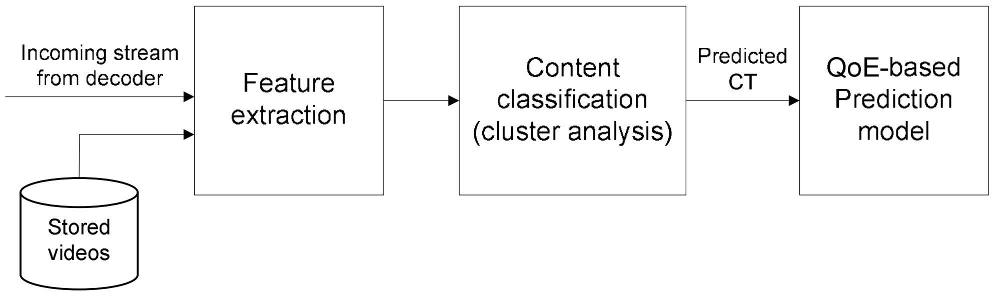


Figure 5.12. Content classification method

Hence, the content classifier takes the content features as input observations, while content category as the output. In this thesis, three distinct content categories are defined as CT=0.1, for slow motion videos e.g. head and shoulder, CT=0.5 for medium movement and CT=0.9 for fast moving sports type of clips. However, they can be continuous from 0 to 1 for larger video clips or movies the input will be segment by segment analysis of the content features extracted. This has been left as future work from this thesis.

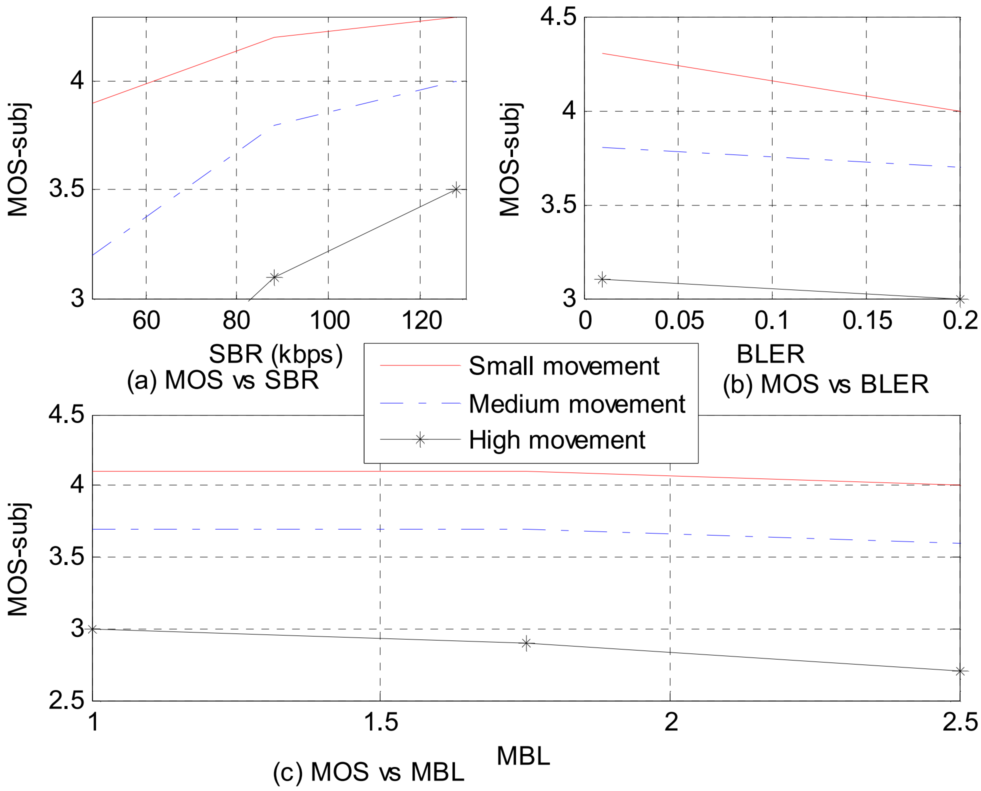


Figure 5.13. Relationships of the QoS parameters with MOS

Fig. 5.13 shows the relationships of video quality (in terms of MOS) to SBR (Fig. 5.13a), BLER (Fig. 5.13b) and MBL (Fig. 5.13c) for all CTs A logarithmic relationship between SBR and MOS was found from Fig. 5.13a (this also re-established the relationship found in Fig. 5.6 earlier. A linear relationship between MOS and CT was found. Also from ANOVA analysis (Table 3.6, 3.7 & 3.8), it was found that the combined impact of SBR and CT is also significant. Therefore the relationships of the encoder related parameters and CT are shown by Equations 5.12-5.14.

$$f_{A1}(SBR) = \gamma_1 ln(SBR) + \gamma_2 \quad (5.12)$$

$$f_{A2}(CT) = \delta_1 CT + \delta_2 \quad (5.13)$$

$$f_{A3}(SBR, CT) = \varepsilon_1 CT * ln(SBR) + \varepsilon_2 \quad (5.14)$$

Relationship of Wireless Access Network (UMTS) Parameters

Similarly, in the access network the following relationship between MOS and BLER and MBL as shown in Figs 5.12b and c was found and can be modelled as polynomial functions. Equations 5.15-5.17 show the relationship of MOS with the access network parameters of BLER and MBL over UMTS access network.

$$f_{P1}(BLER) = \mu_1 + \mu_2(BLER) \tag{5.15}$$

$$f_{P2}(MBL) = \sigma_1 MBL + \sigma_2 \tag{5.16}$$

$$f_{P3}(BLER, MBL) = \eta_1 BLER*MBL + \eta_2 \tag{5.17}$$

5.7 Novel Non-intrusive Video Quality Prediction Models over WLAN

This section describes the model development over WLAN with the simulation set-up described in Section 5.3 earlier. PCA analysis of all the QoS parameters is given followed by MOS prediction.

5.7.1 PCA analysis

Principal Component Analysis (PCA) [83] reduces the dimensionality of the data while retaining as much information as possible. For this reason, PCA was carried out to determine the relationship between MOS and the objective video parameters of SBR, FR and PER. PCA involves calculating eigenvalues and their corresponding eigenvectors of the covariance or correlation matrix. Covariance matrix is used where the same data has the same set of variables and correlation matrix is used in the case where data has a different set of variables. In this thesis, a covariance matrix has been used because the data set is the same.

The PCA was carried out to check the suitability of the objective parameters of SBR, FR and PER for model design. The PCA was performed for the three content types of SM, GW and

RM separately. The variance of the data for the three content types is given in Table 5.5. The first two components account for more than 90% of the variance and hence are sufficient for the modelling of the data. The PCA results are shown in Fig. 5.14.

Table 5.5 Variability of the first two components for all content types

Sequence	Var. of PC1(%)	Var. of PC2(%)
Slight Movement	58	33
Gentle Walking	63	31
Rapid Movement	74	20

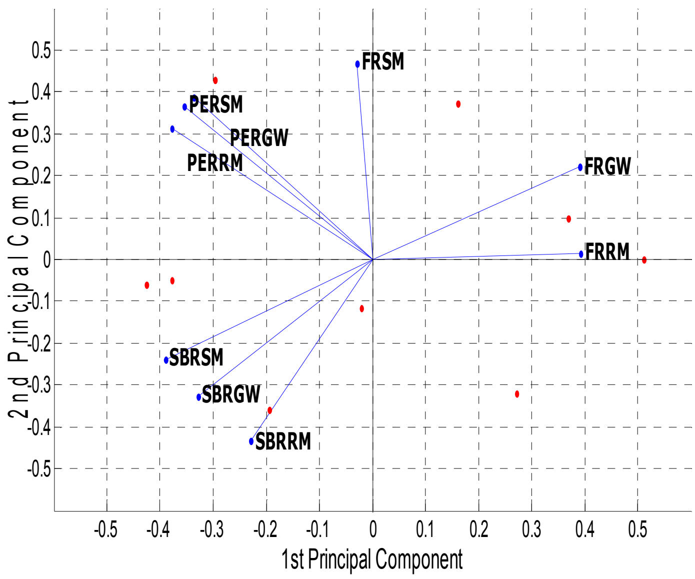


Figure 5.14. PCA results for all content types

The PCA results from Fig. 5.14 show the influence of the chosen parameters (SBR, FR and PER) on the data set for the three content types of SM, GW and RM. In Fig. 5.14 the horizontal axis represents the first principal component (PC1) and the vertical axis represents the second principal component (PC2). Each of the objective parameters (e.g. FRGW, etc) are represented by a vector.

5.7.2 MOS prediction

The proposed model is based on three objective parameters (SBR, FR and PER) for each content type as given by Equation (5.18):

MOS = f(SBR, FR, Content type, PER) *(5.18)*

The prediction model for video quality evaluation is given by a rational model which takes into account the relationships found earlier from Equations (5.1) to (5.6) and is given by Equation (5.19).

$$MOS = \frac{\alpha + \beta FR + \gamma \ln(SBR)}{1 + \mu(PER) + \sigma(PER)^2} \qquad (5.19)$$

The metric coefficients were obtained by a linear regression of the proposed model with the training set (MOS values obtained by objective evaluation given in Table 5.1). The coefficients for all three content types are given in Table 5.6 for validation dataset.

Table 5.6 Coefficients of metric models for all content types over WLAN

Coefficients	SM	GW	RM	All Contents
α	4.5796	3.4757	3.0946	5.1676
β	- 0.0065	0.0022	- 0.0065	-1.4649e-014
δ	n/a	n/a	n/a	-2.4719
γ	0.0573	0.0407	0.1464	-0.0334
μ	2.2073	2.4984	10.0437	3.2078
σ	7.1773	- 3.7433	0.6865	-2.6636

The proposed metric has different coefficient values for the three different content types because spatial and temporal sequence characteristics of the sequences are significantly different. The model's prediction performance is given in terms of the correlation coefficient R^2 (indicates the goodness of fit) and the RMSE (Root Mean Squared Error) and is summarized in Table 5.7.

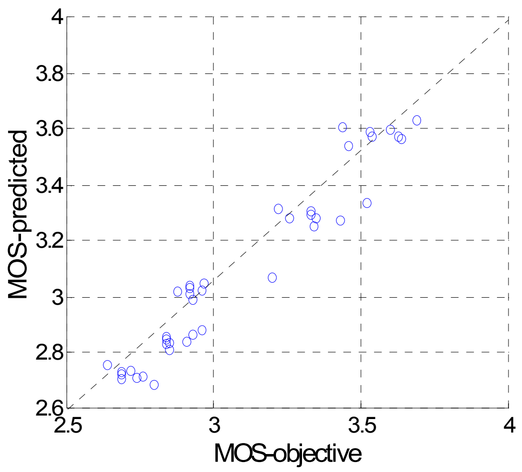
Table 5.7 Metric performance by correlation coefficient and RMSE (WLAN)

Content type	SM	GW	RM	All contents
Correlation coefficient	79.9%	93.36%	91.7%	86.51%
RMSE	0.2919	0.08146	0.2332	0.3366

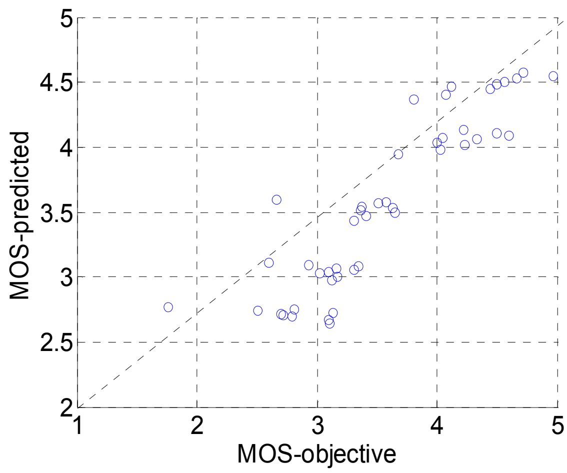
The performance of the video quality prediction obtained by the metric compared to the video quality data obtained objectively using NS2 [43] for content type of 'gentle walking' is shown in Fig. 5.14a. Slightly better correlation was achieved for 'gentle walking' compared to the other two content types of ‘slight movement’ and ‘rapid movement’. It was also observed that video clips in ‘rapid movement’ are very sensitive to packet loss. The quality degrades rapidly compared to the other two categories as packet loss is introduced. Whereas, for ‘slight movement’ the video quality was still acceptable (MOS>3.5) for packet losses of up to 20%.

Further, Eq. 5.20 has been modified to take into account all contents. Therefore, content types then become an input to the model. Equation 5.20 represents the revised model equation which takes CT as an input.

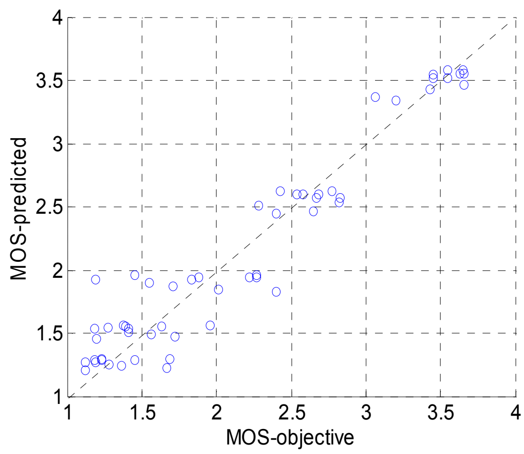
$$MOS = \frac{\alpha + \beta e^{FR} + \gamma \ln(SBR) + \delta CT}{1 + \mu(PER) + \sigma(PER)^2} \qquad (5.20)$$



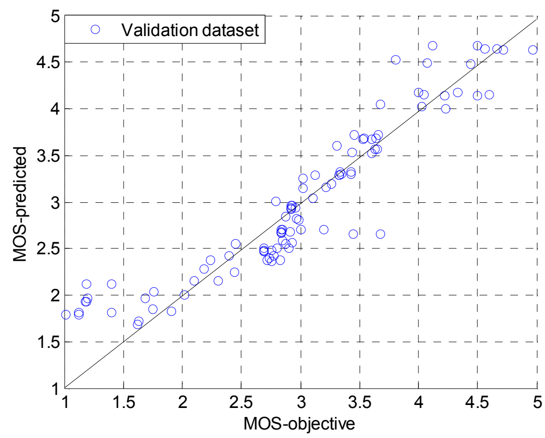
(a) Predicted vs. objective MOS for ‘SM’



(b) Predicted vs. objective MOS for ‘GW’



(c) Predicted vs. objective MOS for ‘RM



(d) Predicted vs. objective MOS for all contents

Figure 5.15. Predicted vs objective MOS over WLAN

Figure 5.15a-c shows model prediction for the three content types, whereas, Fig. 5.15d shows the model prediction using Equation 5.20, where the content type is an input to the prediction model. The performance of the model in terms of correlation coefficient is around 87% as shown in Table 5.7. Therefore, this shows a good estimation of visual quality from the proposed model.

5.8 Novel Non-intrusive Video Quality Prediction Models over UMTS

This section shows the model development over UMTS networks. The simulation set-up is shown in Section 5.3. The models shown have been developed with datasets obtained from NS2 and OPNET simulation.

As it was concluded from earlier results that PSNR is not a good reflector of visual quality, subjective tests were carried out based on existing VoIP tests [59]. Section 5.4 earlier described the subjective test set-up. The models over UMTS are then developed using subjective data.

<u>MOS prediction</u>

Once the relationship of the individual QoS parameters on MOS was established, nonlinear regression analysis was carried out in MATLAB®. The relationships found in Equations (5.12) to (5.17) have been used in the MOS prediction. The following nonlinear Equation was obtained given in Eq. (5.21) with a reasonable goodness of fit:

$$MOS = \alpha + \frac{\beta e^{FR} + \gamma ln(SBR) + CT(\delta + \epsilon ln(SBR))}{1 + (\mu(BLER) + \sigma(BLER)^2)\eta MBL} \tag{5.21}$$

Table 5.8 gives the coefficient of the model given by Eq. (5.21). The model coefficients are given for the datasets generated using NS2 and OPNET simulation platforms.

Table 5.8 Coefficients of metric model for all contents over UMTS

Coefficient	NS2(validation)	OPNET(validation)
α	4.2859	4.3911
β	-3.9832e-009	3.9544e-08
γ	0.1090	0.0447
δ	-1.4684	8.8501
ε	-0.0456	-2.1381
μ	-2.6315	-0.3631
σ	2.9220	-10.1175
η	n/a	0.3442
Corr Coeff	86.19%	86.52%
RMSE	0.2544	0.355

The simulation set-up over NS2 takes MBL = 1 only, assuming a random uniform error model. However, it can easily be extended to take variable MBLs. Table 5.8 also outlines the correlation coefficient and RMSE of the models proposed. From the correlation coefficient – both models give a correlation coefficient of 86%, very little difference in datasets were found over the simulation systems. It was concluded that either simulation system (OPNET/NS2) can be used to simulate UMTS network conditions.

As it was pointed out in Fig. 5.10 that PSNR is not a good reflector of visual quality, subjective tests were carried out over UMTS networks only due to cost limitations according to Section 5.4.

Prior to model fitting, a four-way ANOVA analysis was carried out on the MOS dataset to further confirm the coefficients for model fitting. Table 5.9 show the 4-way ANOVA analysis and shows the CT and SBR have the highest impact followed by BLER and MBL. This also confirms the ranking of QoS parameters as carried out in Chapter 3, Section 3.8.

For subjective dataset, FR was fixed at 10fps and was not a variable. However, the model can be easily extended to include FR as a variable. For subjective data, Equation (5.21) is therefore reduced to Equation (5.22) below.

$$MOS = \frac{\alpha + \gamma ln(SBR) + CT(\delta + \epsilon ln(SBR))}{1 + \mu(BLER)MBL} \qquad (5.22)$$

Table 5.9 Four-way ANOVA on MOS obtained on Subjective dataset

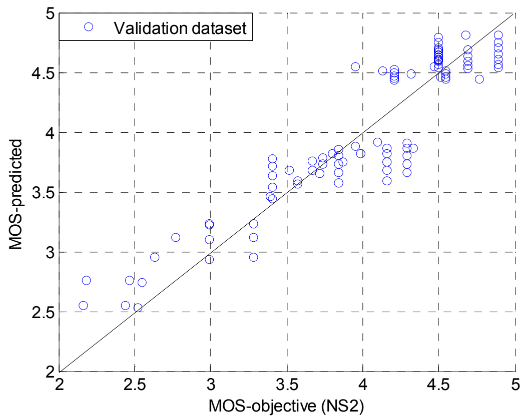
Parameters	Sum of Squares	Degrees of Freedom	Mean Squares	F –statistic	p-value
CT	13.3026	2	6.6513	226.51	0
SBR	2.8159	2	1.40796	47.95	0
BLER	0.2674	1	0.26741	9.11	0.0041
MBL	0.0115	2	0.00574	0.2	0.8231

The model coefficients and correlation coefficient are given in Table 5.10. For the subjective data the coefficients of the model are reduced for better fit and low complexity. This is because subjective tests use a subset of the test conditions used (given by Table 5.3) by simulation due to cost implications of running subjective tests.

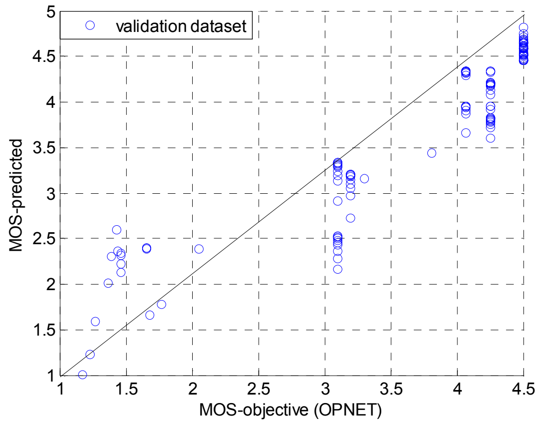
Table 5.10 Metric coefficient (subjective dataset) and performance

Coefficient	Subjective (validation)
α	3.9560
γ	0.0919
δ	-5.8497
ε	0.9843
μ	0.0982
Corr. Coefficient	92.97%
RMSE	0.1103

Figures 5.16a and b show the metric performance against objective MOS data using NS2 and OPNET simulation environments. Equation 5.21 has been used to generate Figs. 5.16a and b. The correlation coefficients for both are around 86% which shows that the model performs well irrespective of the simulation environments.



(a) Predicted vs. objective MOS for NS2



(b) Predicted vs. objective MOS for OPNET

Figure 5.16. Predicted vs Objective MOS over UMTS

Figures 5.17 show model performance against subjective data (PC-based) for the validation dataset. The performance of the metric has been improved to around 93% respectively. Equation 5.22 has been used to generate Fig. 5.17.

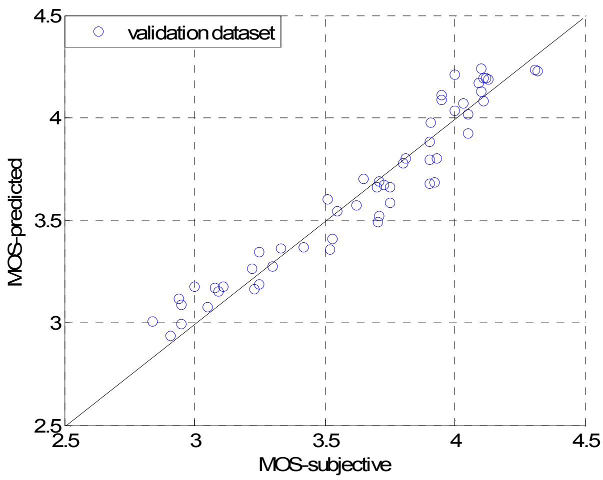


Figure 5.17. Predicted vs Subjective MOS over UMTS (validation dataset)

5.9 Model comparison and validation with external databases

The proposed models were validated against the (1) objective metrics of PSNR and SSIM, (2) subjective data taken from [15] (encoder based parameters of SBR and FR and CT) and from [79] (both encoder based and access network based) for H.264 video and (3) subjective data taken from handset-based tests carried out at EHU, Spain. The results were not compared against VQM data [2],[122] as it is limited to MPEG4 and H263 codecs. Figure 5.18 shows the end-to-end correlation analysis of the proposed model with objective metrics of PSNR and SSIM and with the subjective tests.

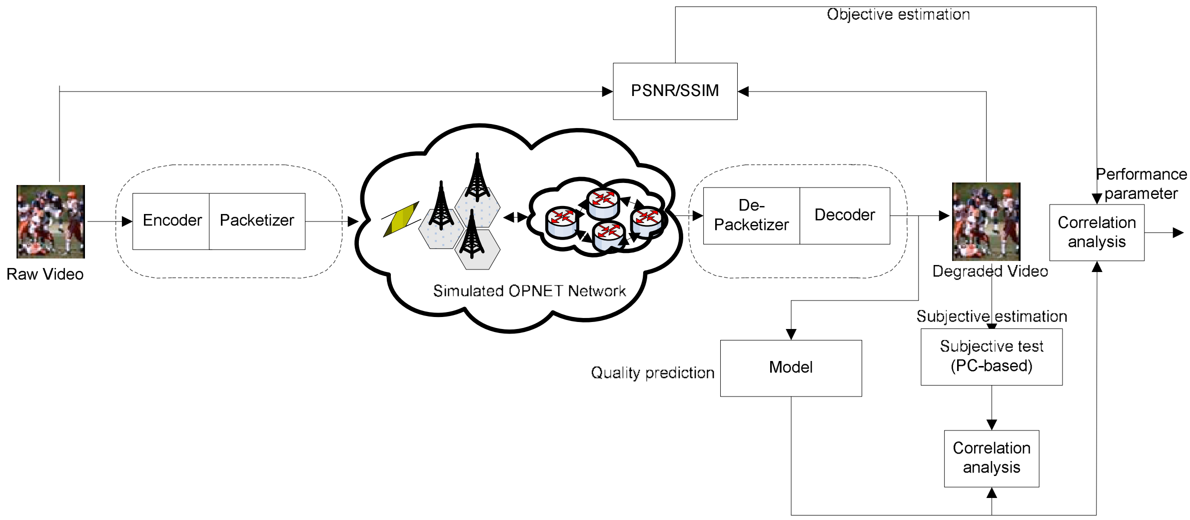


Figure 5.18 Correlation of subjective vs objective MOS with proposed model

Figs. 5.19a and b shows the scatter plots of different metrics versus MOS. The proposed video quality metric was compared with the most widely used Peak-Signal-to-Noise-Ratio (PSNR) metric and with Structural Similarity Index Metric SSIM [3]. The corresponding R^2 values of the fittings are 93% (from our proposed model), 66.73% (from PSNR) and 76.41% (from SSIM) which suggests that the proposed metric is more correlated with the subjective MOS values.

Fig. 5.20a and b show the model correlation against external data from [15] and [79]. The data from [15] is for H.264 QCIF videos with application layer parameters of frame rate, SBR and CT. The data from [15] is for H.264 QCIF videos with encoder parameters of FR and SBR. Encoder only parameters were chosen in Eq. (5.22) to show that if network losses were taken as zero (BLER=0 in Eq.(5.22)), the model gives a correlation of ~78%. Equation (5.22) reduces to Eq. (5.23) as follows:

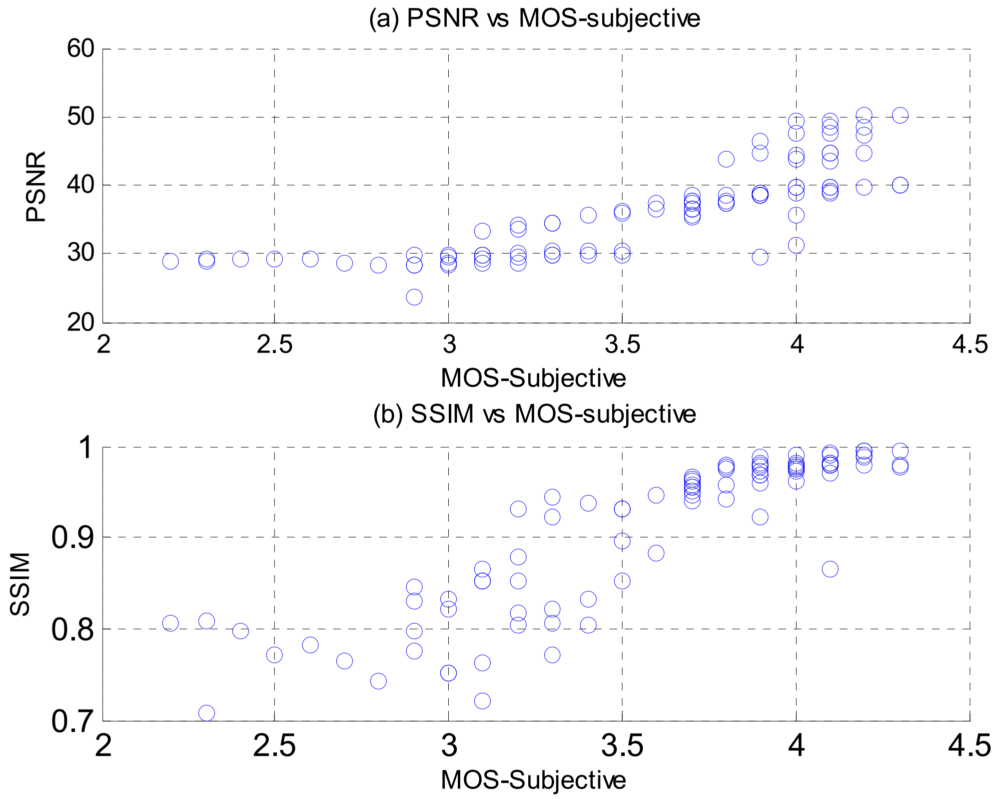


Figure 5.19. Scatter plots of PSNR (a) and SSIM (b) versus MOS-subjective

$$MOS = \alpha + \beta.FR + \gamma.\ln(SBR) + CT(\delta + \varepsilon.\ln(SBR)) \qquad (5.23)$$

Whereas, data in Fig 5.20b is from [18],[79] which is H.264 videos of size 768 x 480 with high SBRs and packet losses. The frame rate was fixed at 30fps. A correlation coefficient of ~ 78% for data from [15] and ~74% for data from [79] was achieved. Even though the model was trained with QCIF videos of low bitrates, it works with larger size videos with very high bitrates. Fig. 5.20b shows little variation when DMOS is between 25-40. The variation in content types in the LIVE dataset is very little. The content classifier classified most of the content as slow to medium activity and hence this caused the lack of variation at low DMOS values. In addition, LIVE wireless video database considered videos encoded at maximum frame rates, whereas the model is derived from subjective data where test conditions included only one fixed frame rate (10fps). However, literature has provided evidence that subjective quality varies with frame rate and that it is a result of interaction between temporal and spatial dimensions. As the model does not take that aspect into account, is probably a reason why it does not perform so well on the LIVE wireless video database.

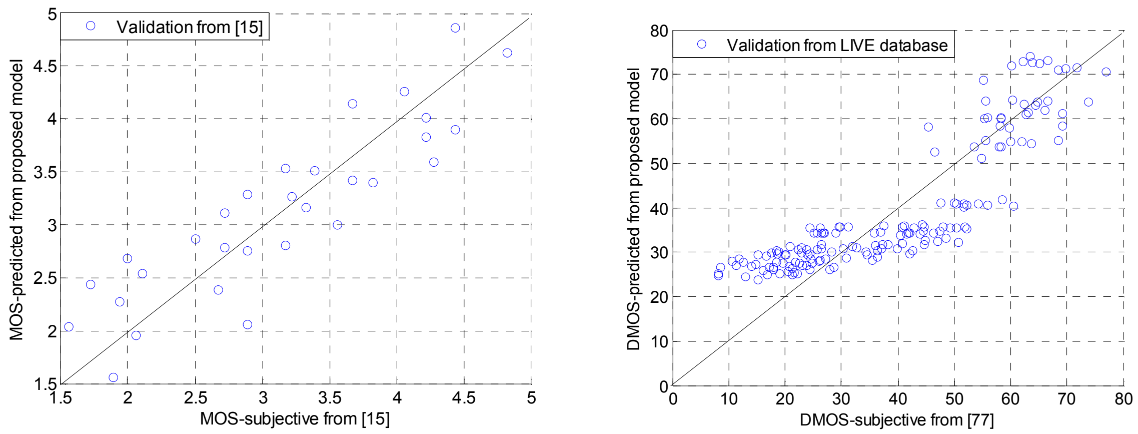
Finally, the model was validated against the datasets used at EHU, Spain with the handset-based subjective tests. A correlation coefficient of ~ 69% was achieved.

Table 5.11 summarizes the correlation coefficient and root mean squared error of the model against the external databases and PSNR and SSIM.

Table 5.11 Model validation correlation coefficients

	R^2	RMSE
Validation against PSNR	66.73%	3.793
Validation against SSIM	76.41%	0.0407
Validation against [15]	77.9%	0.3812
Validation against [77]	74.07%	7.333
Validation against EHU (Handset-based MOS)	68.55%	0.4432

Therefore, it can be concluded that the proposed video quality model can serve as a simple yet accurate metric to measure perceptual video quality and hence an input in to the proposed QoS-driven adaptation scheme (See Chapter 7).



(a) Validation of proposed model from subjective data in [15]

(b) Validation of proposed model from LIVE subjective data in [79].

Figure 5.20. External validation of model

The model performance from the handset-based dataset is given in Fig. 5.21. It performed slightly lower than with the other datasets. This was because of the very low MOS scores given to the fast moving content – mainly because it was viewed on a small screen and so had an impact. Fig. 5.21 shows the performance of the model against subjective data collected at EHU using mobile handset.

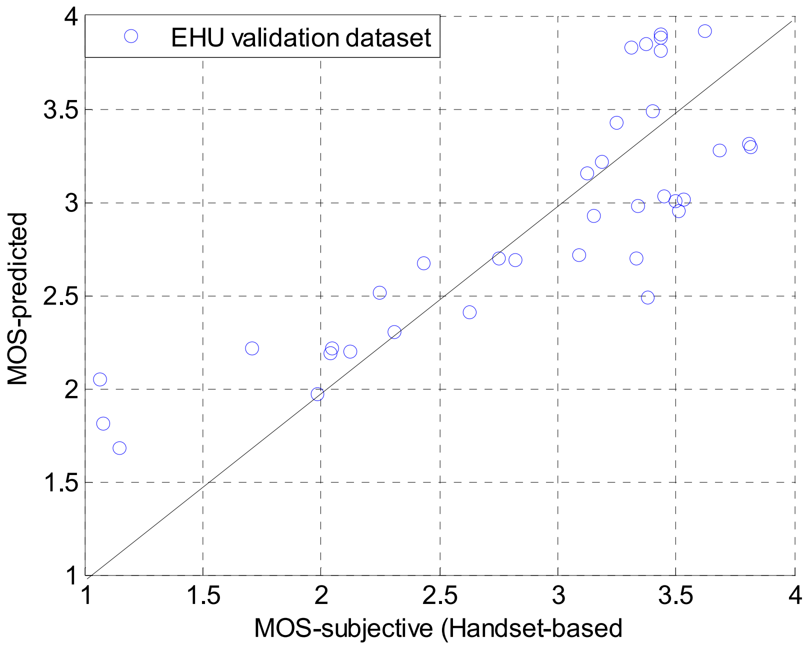


Figure 5.21. Validation of model from Handset-based (EHU) dataset

5.10 Summary

In this chapter, novel non-intrusive video quality prediction models over wireless access networks of WLAN and UMTS are presented using non-linear regression modelling. The prediction is based on both objective and subjective datasets. Two codecs were used. H.264 for video transmitted over UMTS and MPEG4 over WLAN. The proposed models were validated against external databases and performed well overall in terms of correlation coefficient. The external database validation also highlighted potential areas where the proposed models did not perform as well. Live dataset had very little content variety. This combined with the frame rate variation resulted in the model performing not as well in some parts. Also the verification with the handset dataset was interesting. As it was concluded that the end device have an impact on video quality prediction. Thus, for future work all subjective data should be collected on the end device application for modeling.

The application of these models in video quality optimization and QoS control are given in Chapters 7.

Chapter 6

Neural Network-based Models for Non-intrusive Video Quality Prediction over WLAN and UMTS

6.1 Introduction

In Chapter 5, the proposed non-intrusive video quality prediction models have been presented based on non-linear regression analysis. Regression models proposed in Chapter 5 are static and can only adapt to change in access network conditions is they are used in an adaptation mechanism. However, the advantage of using artificial neural network models to predict video quality non-intrusively is that the neural network can adapt to changing network conditions due to their ability to learn

There is limited work in literature where ANN have been used in assessing the video quality from both network and application based parameters. The related work from literature is given in the next section under "Related Works". There are many parameters that affect video quality and their joint affect is not clear, and their relationships are thought to be non-linear. Artificial Neural Networks (ANNs) can be used to learn this non-linear relationship which mimics human perception of video quality. In this Chapter learning models based on ANFIS (Adaptive Neural Fuzzy Inference System) are developed to predict the visual quality in terms of the Mean Opinion Score (MOS) for all contents over access networks of UMTS and WLAN. ANFIS is well suited for video quality prediction over error prone and bandwidth restricted wireless access networks as it combines the advantages of neural networks and fuzzy systems. ANFIS was specifically chosen over a back propagation neural network because it includes fuzzy inference system which takes human reasoning (linguistic information) into account. Also the input data is not needed to be normalized. The back propagation in ANFIS optimizes the fuzzy logic. In addition work presented in [14] first used artificial neural networks in video quality assessment and concluded that random neural networks out performed feed forward artificial neural networks. As part of their future work they recommended use of better techniques within neural networks would enhance video quality assessment. Keeping that in mind, in this work, it was felt that the contribution in video quality assessment would be furthered if the same neural networks models were not repeated. Instead use an enhancement of neural networks that include fuzzy logic and hence, improves the results in video quality assessment and prediction.

The structure of the chapter is as follows. Section 6.2 gives the related work from literature. The background to the ANFIS-based ANN is given in Section 6.3. Section 6.4 outlines the experimental set-up over WLAN and UMTS. Section 6.5 describes the procedure to develop the ANFIS-based models. The ANFIS-based models over WLAN are presented in Section 6.6. Section 6.7 presents the models over UMTS. Model validation with external LIVE database is given in Section 6.8. Section 6.9 summarizes the Chapter.

6.2 Related work

There are many parameters that affect video quality and their combined affect is not clear, and their relationships are thought to be non-linear. Artificial Neural Networks (ANNs) can be used to learn this non-linear relationship which mimics human perception of video quality.

ANN has been widely used in assessing the video quality from both network and application based parameters. In [14],[13],[123] the authors have developed neural-network models to predict video quality based on application and network parameters. The work was based on video subjective tests to form training and testing datasets. Further, different video contents have not been considered in developing neural network models and their work is only limited in fixed IP networks. In [124] PSQA (Psuedo-Subjective Quality Assessment) metric based on random neural networks to evaluate the quality of multimedia communication over home networks has been presented. Work presented in [125] use a set of parameters associated with the encoder (sender bitrates and frame rates) and content types and have neural networks for video quality prediction. In [126] raw video features extracted from H.264 encoded videos has been used to train the neural network for video quality prediction. Similarly, in [127] neural networks based on raw video features for MPEG2 video sequences has been presented. Video content features have been used in [128] to train the neural network for reduced reference video quality assessment. Several neural network architectures have been compared in [129] to predict MPEG4 traffic pattern.

The choice of parameters used in literature is either raw features or encoder based distortion to train the neural network. Further, different video contents have not been considered in developing neural network models and their work is only limited in fixed IP networks. The choice of parameters is crucial to video quality prediction as neural networks have been successfully used in literature for video quality assessment and prediction.

6.3 Background to ANFIS-based ANN

ANFIS uses a hybrid learning procedure and can build up an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and specified input-output data pairs. A two input ANFIS architecture [130] as shown in Fig. 6.1 is an adaptive multilayer feed forward network in which each node performs a particular function on incoming signals as well as a set of parameters pertaining to this node.

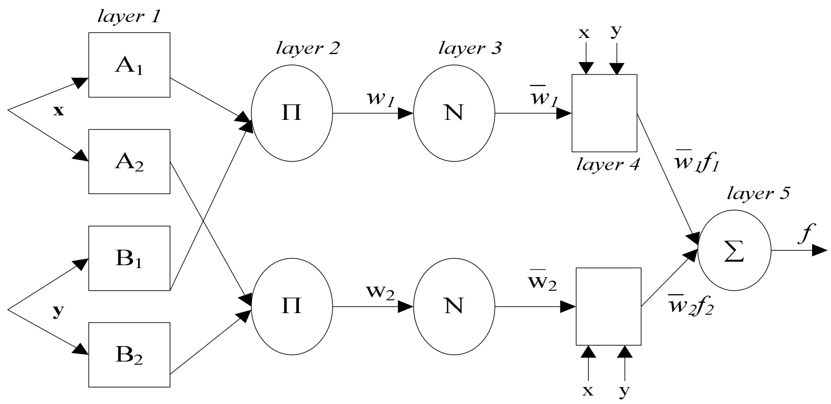


Figure 6.1. ANFIS architecture [130]

The ANFIS construction is given in Fig. 6.1. It consists of five layers. The layers are called a fuzzy layer, a product layer, a normalized layer, a defuzzy layer and a total output layer. There are two inputs in the ANFIS structure shown in Fig. 6.1 as x and y. The output is *f*. For a first-order Sugeno fuzzy model, a typical rule set with two fuzzy if–then rules can be expressed as:

Rule 1: If x is A_1 and (y is B_1) then $f_1=p_1x+q_1y+r_1$

Rule 2: If x is A_2 and (y is B_2) then $f_2=p_2x+q_2y+r_2$

where p_1 , p_2 , q_1 , q_2 , r_1 and r_2 are linear parameters, and A_1, A_2, B_1 and B_2 are nonlinear parameters. The output f is given by Eq. (6.1).

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} \tag{6.1}$$

The fourth layer is the defuzzy layer, whose nodes are adaptive. The output Equation is $w_i(p_ix+q_iy+r_i)$, where p_i, q_i and r_i denote the linear parameters or so-called consequent parameters of the node. The defuzzy relationship between the input and output of this layer can be defined as given by Eq. (6.2):

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i\,(p_i x + q_i y + r_i) \tag{6.2}$$

where $O_{4,i}$, denotes the Layer 4 output. The fifth layer is the total output layer, whose node is labelled $\sum$. The output of this layer is the total of the input signals, which represents the shift decision result. The results can be written as given by Eq. (6.3):

$$O_{5,i} = overall\ output = \sum_i \bar{w}_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad (6.3)$$

where $O_{5,i}$ denotes the Layer 5 output [130].

6.4 Experimental set-up for the ANFIS-based Models to Predict Video Quality

This section outlines the experimental set-up used over WLAN and UMTS for the development of the ANFIS-based models.

6.4.1 Experimental set-up over WLAN

The datasets and simulation set-up are the same as given in Section 5.3.1 over WLAN.

6.4.2 Experimental set-up over UMTS

The datasets and the simulation set-up are the same as given in Section 5.3.2 over UMTS. Further, the subjective datasets and experimental set-up are the same as given in Section 5.4 over UMTS.

6.5 Procedure to develop the ANFIS-based models

This section describes the procedure to develop the ANFIS-based models, their membership functions and the MATLAB functions used.

6.5.1 Block diagram of the ANFIS-based models

The block diagram of the video quality prediction system is depicted in Fig. 6.2. The ANFIS-based prediction model is at the receiver side to predict video quality non-intrusively.

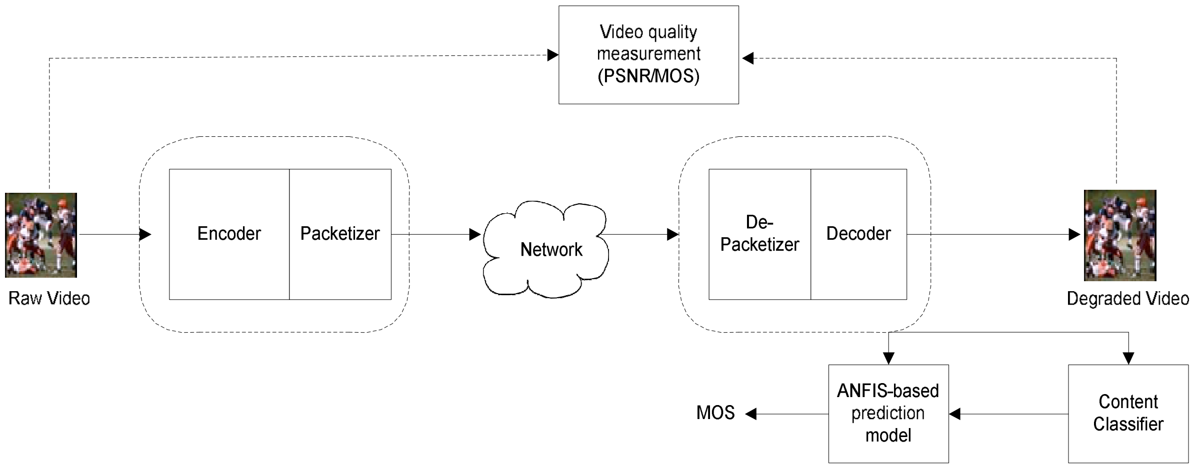


Figure 6.2. End-to-end diagram of video quality prediction with ANFIS-based model

The aim is to develop ANFIS-based learning models to predict perceived video quality for three distinct content types from a combination of parameters related to the encoder and the wireless access network. The functional block of the proposed model is shown in Fig.6.3.

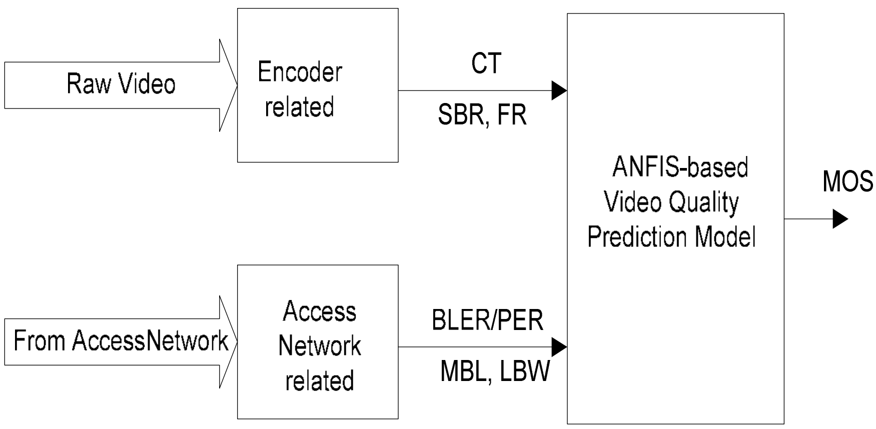


Figure 6.3. Block diagram of the ANFIS-based prediction model

For the tests three different video sequences representing slow moving content to fast moving content were used. The video sequences were of QCIF resolution (176x144) and encoded in H.264 format with an open source JM software [46] encoder/decoder for UMTS access network and in MPEG4 [45] for WLAN access network. The three video clips were transmitted over simulated UMTS and WLAN access networks using NS2 simulator. The encoder related parameters are Frame Rate (FR) and Sender Bitrate (SBR). The access network parameters are Block Error Rate (BLER) and Link Bandwidth (LBW) for UMTS and Packet Error Rate (PER) and LBW for WLAN for different types of content.

6.5.2 Training and validation of ANFIS-based models

For ANNs, it is not a challenge to predict patterns existing on a sequence with which they were trained. The real challenge is to predict sequences that the network did not use for training. However, the part of the video sequence to be used for training should be 'rich enough' to equip the network with enough power to extrapolate patterns that may exist in other sequences. Three different content types representing different scenarios from slow movement to fast moving sports clips are chosen for training purposes and three different video clips for validation purposes. The ANFIS-based ANN model were trained with the three distinct content types of 'Akiyo', 'Foreman' and 'Stefan' (see Tables 5.1, 5.2 & 5.3) and validated by three different content types of 'Suzie', 'Carphone' and Football' in the corresponding content categories. The data selected for validation was one third that of testing with different parameter values to that given in Tables 5.1, 5.2 & 5.3. In total there were around 450 encoded test sequences for training and 210 encoded test sequences for validation over NS2 over WLAN and UMTS for each model from Tables 5.1 and 5.2. There were 567 sequences for training and 224 for validation over OPNET over UMTS networks. There were 81 test sequences for training and 54 for validation from Table 5.3 (subjective data over UMTS). This is the same as used in Chapter 5.

MATLAB Functions used

The ANFIS functions used were GENFIS1, ANFIS and EVALFIS from MATLAB. More details on these functions can be found from [131]. There are mainly three functions used by the ANFIS neural network as GENFIS1, ANFIS and EVALFIS. It is the GENFIS1 function that generates the Sugeno type FIS which is used by ANFIS for training of the data. The EVALFIS function then performs the fuzzy inference calculations.

Fig. 6.4 summarizes the steps in a flowchart.

An example of the dataset used for model training is shown in Table 6.1.

Table 6.1 Variables used in ANFIS database generation

FR	CT	SBR	LBW	BLER/PER	MBL	MOS
10	0.1	18	32	0.01	1	3.73
10	0.1	18	32	0.05	1	3.56
10	0.1	18	32	0.10	1	2.95
10	0.1	18	32	0.15	1	2.94
10	0.1	18	32	0.20	1	2.92
10	0.1	18	64	0.01	1	3.65
…	…	…	…	….	..	….

Note:

The inputs of FR, CT, SBR, LBW and PER were used to train the model over WLAN.

The inputs of FR, CT, SBR, LBW and BLER were used to train the model over UMTS using NS2 simulation.

The inputs of FR, CT, SBR, BLER and MBL were used to train the model over UMTS using OPNET.

The inputs of CT, SBR, BLER and MBL were used to train the model over UMTS using subjective tests (PC-based).

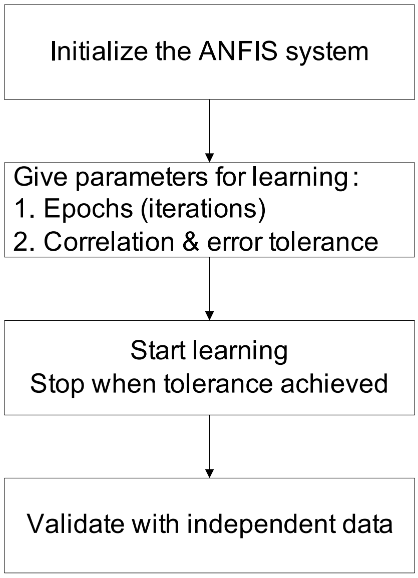


Figure 6.4. Flow chart of ANFIS system

6.5.3 ANFIS Architecture

The corresponding equivalent ANFIS architecture for the two learning models developed (5 inputs respectively) over WLAN and UMTS access networks is shown in Fig. 6.5.

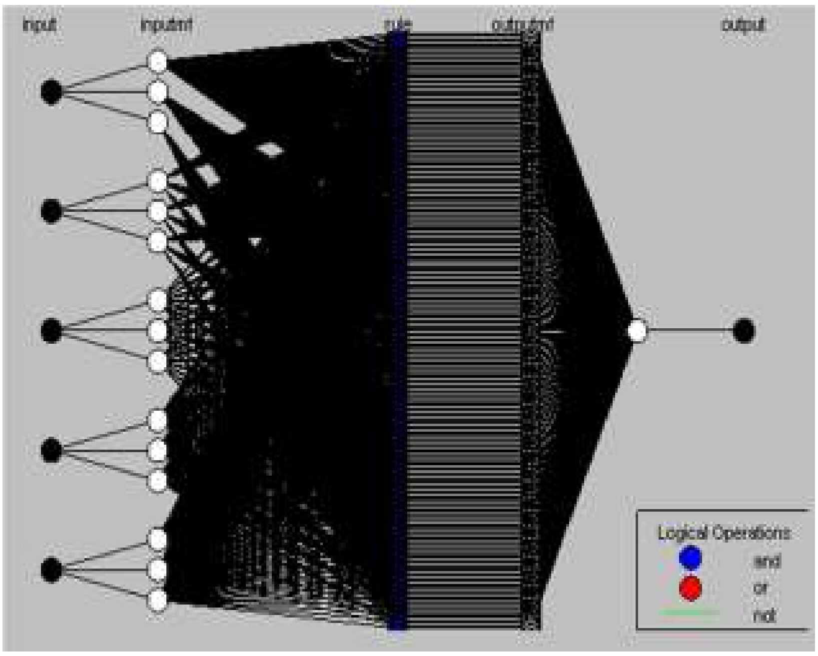


Figure 6.5. ANFIS architecture

There are five inputs as CT, FR, SBR, LBW and BLER/PER. Output is the MOS value. The membership functions for the ANFIS learning model given in Figs. 6.6 and 6.7 respectively as *inputmf* for UMTS and WLAN.

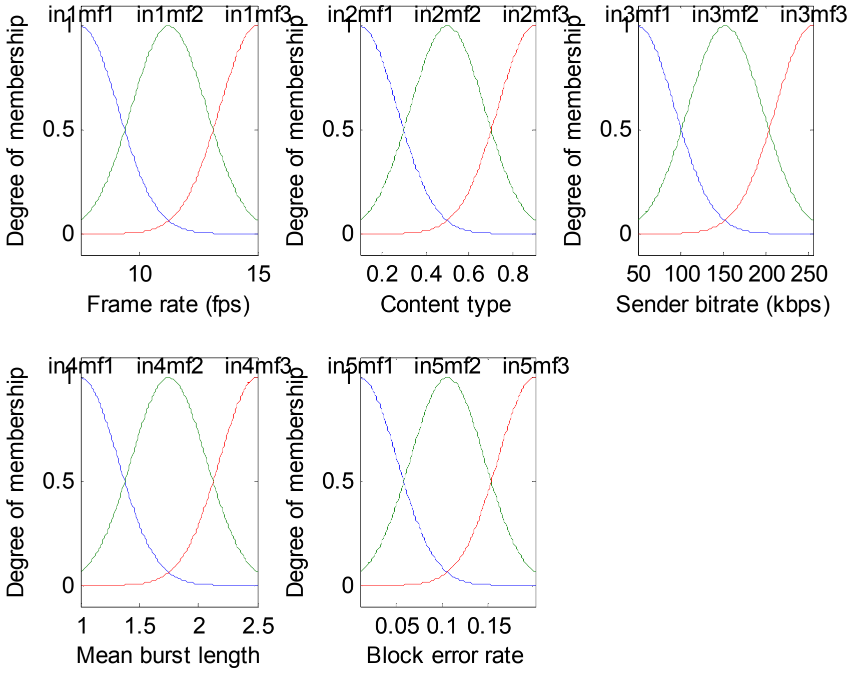


Figure 6.6. Membership functions for 5 inputs over UMTS (OPNET)

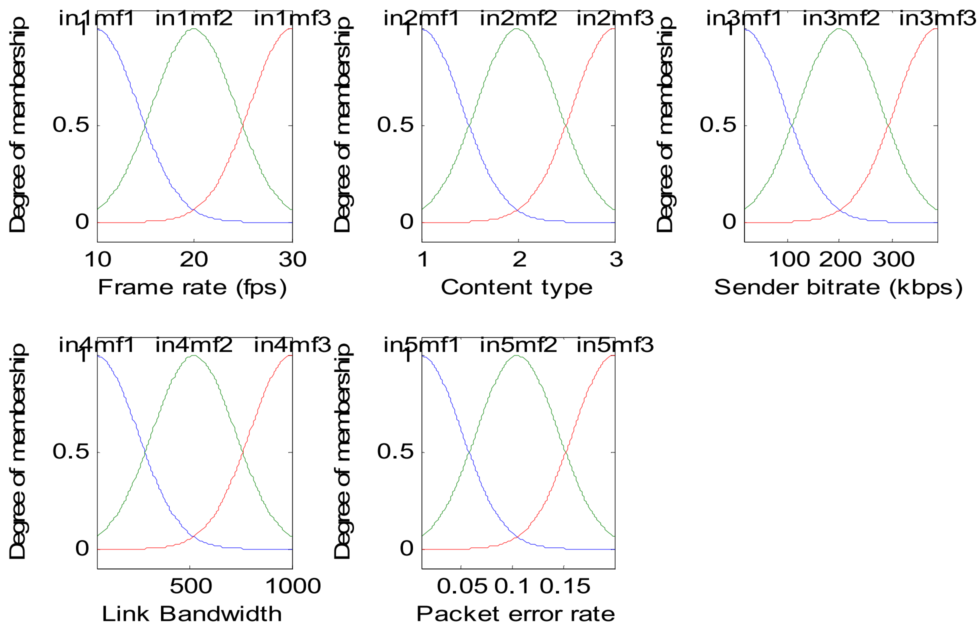


Figure 6.7. Membership functions for 5 inputs over WLAN

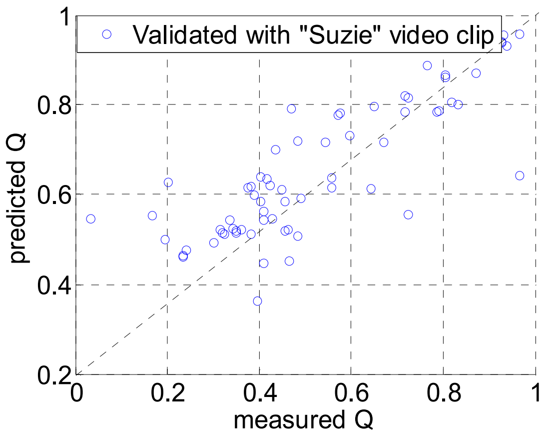
The number of membership function is chosen to be three for all five inputs and their operating range depends on the five inputs. For example for input of SBR the operating range is from [48,256]. Whereas, the CT ranges from [0 1] representing the three different types of content. In the simulation all the Membership Functions (MF) used are generalized bell function defined in [130] as given by Eq. (6.7):

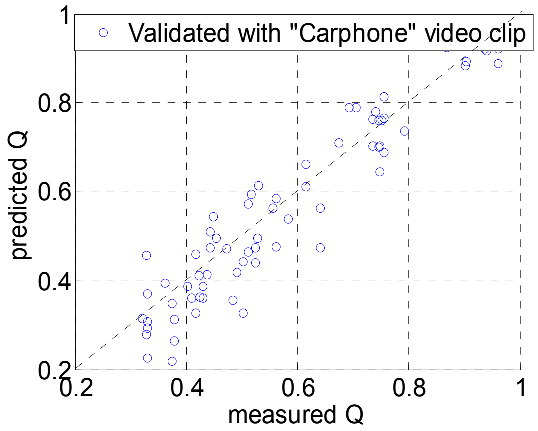
$$\mu_a = \frac{1}{1 + [(\frac{x-c}{a})^2]^b} \qquad (6.7)$$

Which contains the fitting parameters a, b and c. These parameters (a, b, c) are defined such that c finds the centre of the corresponding membership function, the half width is a and the slopes at the crossover points (where MF value is 0.5 in Figs. 6.5 and 6.6.) are controlled by b (together with a).

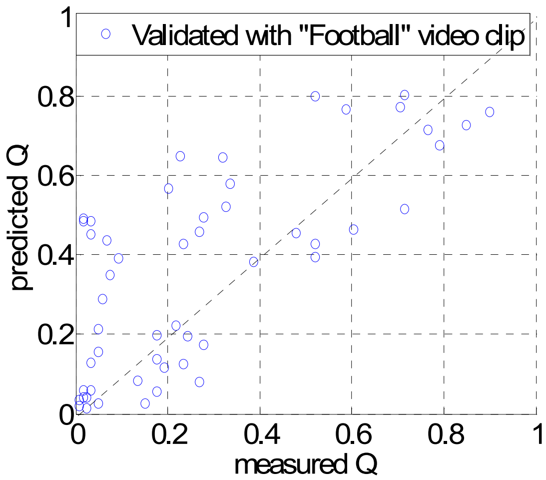
6.6 ANFIS-based Models to Predict Video Quality over WLAN

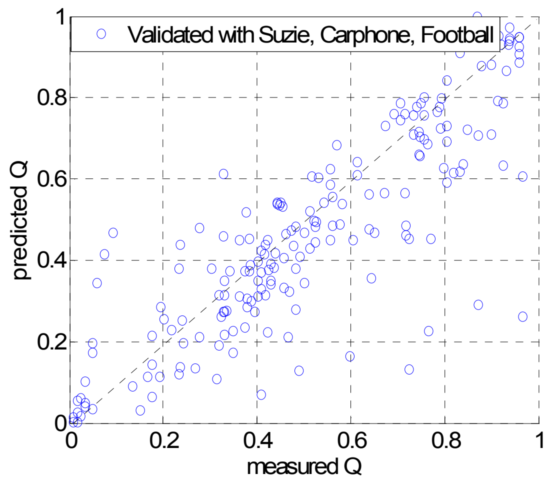
Initially, three ANFIS-based learning models were trained for the three distinct content types and validated them with three different video test sequences in the corresponding content categories. The accuracy of the ANN can be determined by the correlation coefficient and the RMSE of the validation results. For the three content types results were obtained in terms of the MOS values (obtained from PSNR to MOS conversion) and decodable frame rate Q. Further, the three content types were combined and trained as one model with CT as an input in terms of the MOS.





(a) ANN mapping of predicted Q for ‘SM’ (b) ANN mapping of predicted Q for ‘GW’

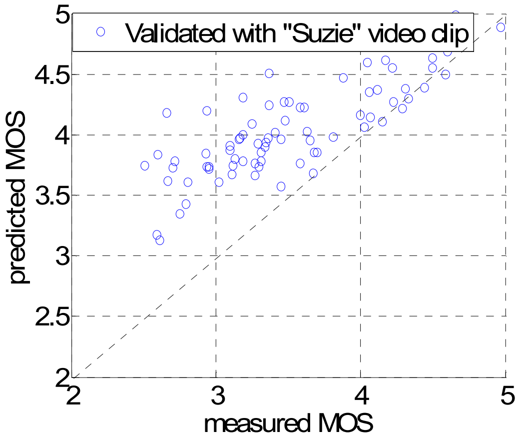




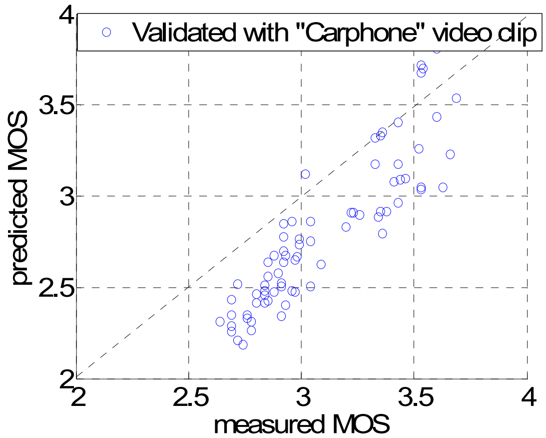
(c) ANN mapping of predicted Q for ‘RM’ (d) ANN mapping of predicted Q for all contents

Figure 6.8. ANN mapping of predicted Q vs measured Q over WLAN

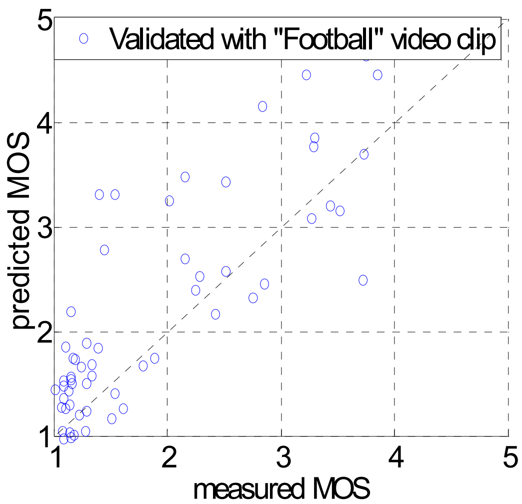
Figs. 6.8a-c gives the predicted mapping of Q vs the measured one for three content types. Fig. 6.8d gives the same for all contents. Table 6.2 summarizes the correlation coefficient and RMSE. It was found that the results for MOS (obtained from PSNR to MOS conversion) were better in terms of correlation coefficient and RMSE. This could be due the fact that Q gave worse results for content type of ‘RM’. It was interesting to use Q and compare it to the widely used PSNR. However, our results show that in terms of reflecting the quality there was very little difference between Q and PSNR. For that reason PSNR was chosen in the next section and obtained MOS from PSNR to MOS conversion given by Evalvid.



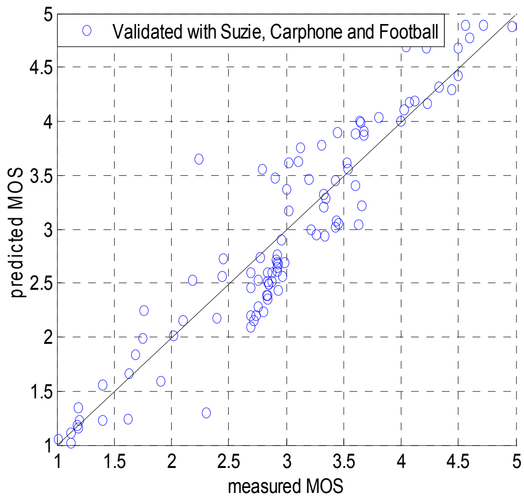
(a) ANN mapping of predicted MOS for 'SM'



(b) ANN mapping of predicted MOS for 'GW'



(c) ANN mapping of predicted MOS 'RM'



(d) ANN mapping of predicted MOS for all contents

Figure 6.9. ANN mapping of objective MOS vs predicted MOS over WLAN

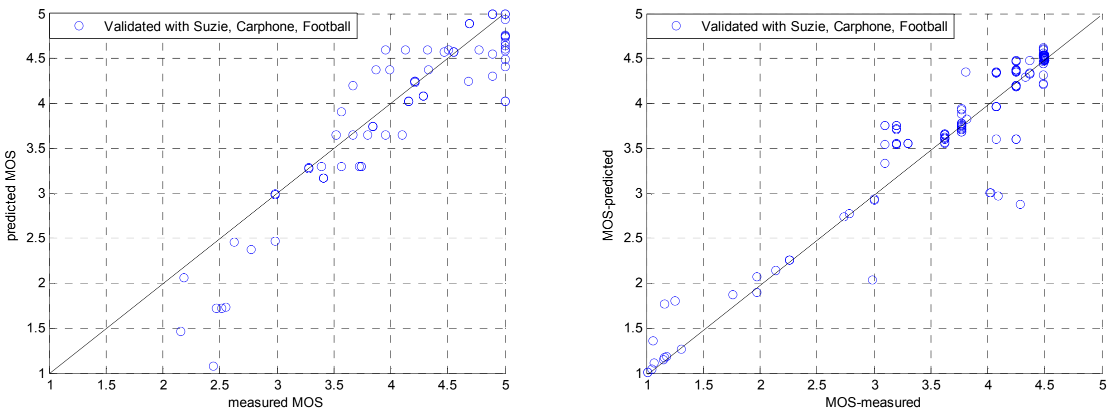
Figs. 6.9a-c gives the results of ANN mapping for MOS for the three content types. Fig. 6.9d gives the same for all types of content. A better correlation coefficient (~90%) was achieved when content type is an input to the ANN. The correlation coefficients and RMSE are summarized in Table 6.2.

Table 6.2 ANFIS prediction performance by correlation coefficient and RMSE (WLAN)

Content type	SM	GW	RW	All Contents
Corr Coef (MOS)	70.07%	80.56%	75.4%	90.42%
RMSE (MOS)	0.1545	0.1846	0.5659	0.31
Corr Coeff (Q)	73.84	92.29%	69.11%	72.05%
RMSE (Q)	0.08813	0.06234	0.2181	0.136

6.7 ANFIS-based Models to Predict Video Quality over UMTS

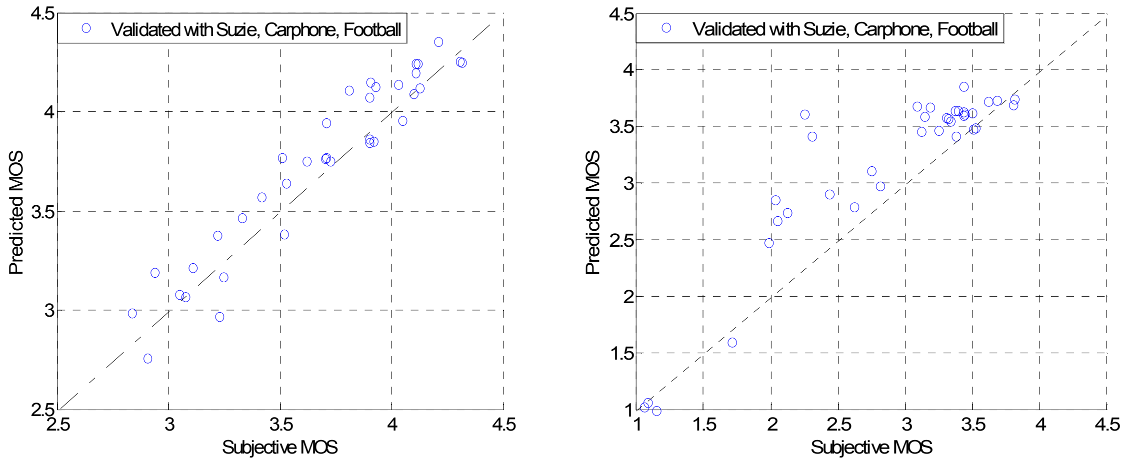
The models over UMTS are presented with simulation platform of NS2 and OPNET. Further, subjective tests were carried out and the models are trained with subjective datasets (Section 5.4).



(a) NS2-based (b) OPNET-based

Figure 6.10. ANN mapping of objective MOS vs predicted MOS over UMTS

Figs. 6.10a and b show the ANN mapping of predicted MOS vs objective MOS over NS2 and OPNET respectively. Table 6.3 summarises the correlation coefficient and RMSE of the ANN mapping. The difference between NS2 and OPNET is non-existent. This shows that data collected over both simulation platforms model UMTS access error well and the ANN model works well for both simulation platforms.



(a) Subjective-PC-based (b) Subjective-Handset-based

Figure 6.11. ANN mapping of subjective MOS vs predicted MOS over UMTS

Figs. 6.11a and b show the ANN mapping of predicted MOS vs subjective MOS data collected over two devices – PC and handset. PC-based results are better than those of the handset based. This could be due to the reason that subjects gave drastically low scores when blocks were lost and due to the small screen. This is looked in more detail in the sub-section 5.4.4. Table 6.3 summarises the correlation coefficient and RMSE.

Table 6.3 ANFIS prediction performance by correlation coefficient and RMSE (UMTS)

Content type	NS2	OPNET	Subjective (PC)	Subjective (Handset)
Corr Coef	86.91%	87.17%	91.15%	82.62%
RMSE	0.3247	0.2812	0.1332	0.3334

6.8 Model comparison and validation with external LIVE database

In this section the models are compared over NS2 and OPNET for objective MOS dataset and PC and handset for subjective dataset. Further, external database of LIVE [79] is used to validate the ANFIS based models. Two thirds of LIVE data is used for model training and one third of LIVE data is used for model validation.

6.8.1 Comparison of the models

The ANFIS-based WLAN model outperformed the model over UMTS in terms of prediction accuracy for objective MOS data. However, the difference in performance was not massive. The performance of the ANFIS-based model over UMTS (both NS2 and OPNET) was slightly worse because the error correlation properties of the link layer do not have an impact on the quality of the streamed video as long as the IP packet error probability remains unchanged. The UMTS model with PC-based subjective data performed very well compared to the one with objective MOS data. The subjective dataset was a subset of the objective dataset due to cost constraints. As a result, the test conditions in subjective data were a lot less compared to those from objective MOS. The handset-based performed did not perform as well as the PC-based model. Nevertheless, the results give an indication that the model performs well in terms of correlation coefficient with the right choice of QoS parameters.

Table 6.3 outlines the correlation coefficient and Root Mean Squared Error (RMSE) of the two models proposed. The PC-based model performed better in terms of the correlation coefficient. It was also found that the range of subjective MOS were dependent on the end device. The results from the ANFIS-based proposed models suggests that the choice of parameters are crucial in achieving good prediction accuracy and confirms the choice of parameters and ANFIS as a prediction tool.

It was found that faster moving content gave low MOS scores over UMTS compared to WLAN. This could be due to the bandwidth restriction over UMTS network for faster moving content types. Also contents with less movement require low sender bitrate to that of higher movement to give acceptable quality.

6.8.2 Validation of models from LIVE external database

Fig. 6.12 show the validation with external LIVE [79] database. LIVE database collected quality ratings in terms of DMOS and is for high sender bitrates.

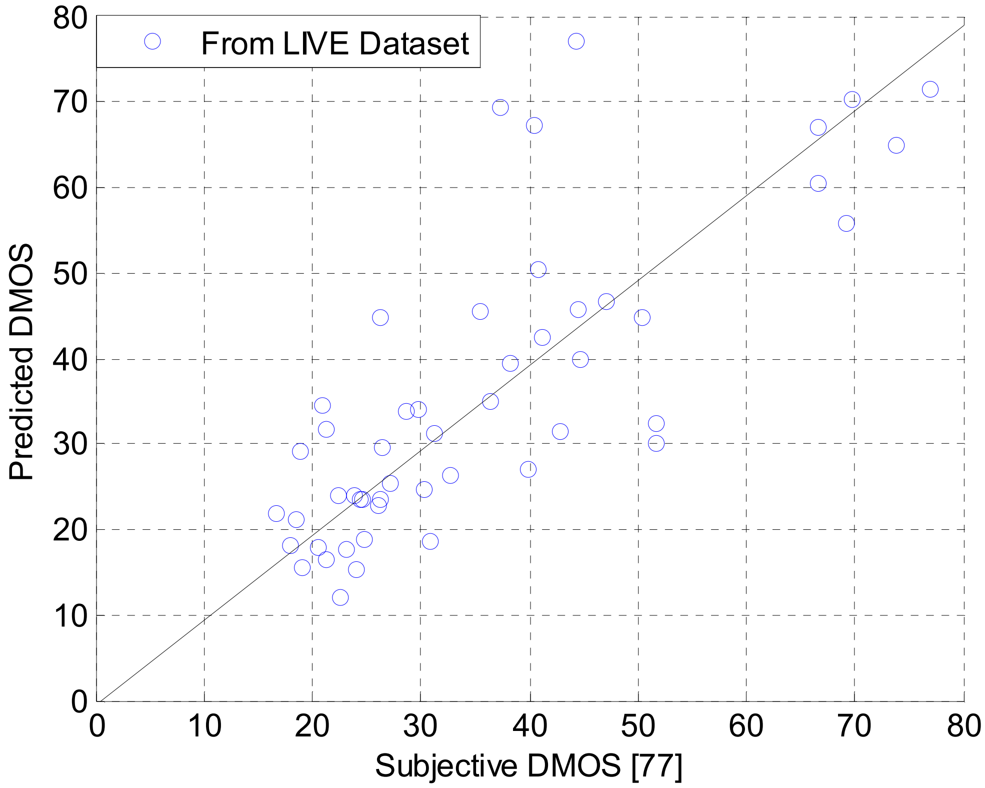


Figure 6.12. ANFIS validation with LIVE dataset

Table 6.4 shows the correlation coefficient with LIVE dataset of the proposed ANFIS model and compares it to that achieved earlier with the regression model. Correlation coefficient of around 72% was achieved. One of the reasons for low correlation coefficient could be the lack of content variety in terms of high movement content in the LIVE dataset. Another one could be the fact that LIVE dataset uses videos at 30fps (maximum frame rates), whereas, the model validated in Fig. 6.12 take 10fps fixed. Comparing the performance of the ANFIS-based model to the regression-based model (Table 6.4), the correlation coefficient is very similar. As subjective quality varies with frame rate and is a result of interaction between temporal and spatial dimensions, the models do not perform so well.

Table 6.4 Comparison of correlation coefficient with regression model for LIVE data

		ANFIS Model	Regression based
LIVE database [79]	Correlation coefficient	71.64%	74.07%
	RMSE	8.568	7.333

6.9 Summary

In this chapter, models based on ANFIS neural network have been developed for predicting video quality in a non-intrusive manner. The models avoid time consuming subjective tests. The neural network database generation and chosen parameters are described. Further the models are compared in terms of correlation coefficient and root mean squared error. The models are then trained with subjective datasets over UMTS networks. Their performance improves with subjective datasets. Finally, external LIVE dataset are used to validate the models and perform very well.

Chapter 7

QoS-driven optimisation and control

7.1 Introduction

In Chapter 5 nonlinear regression models have been developed for predicting video quality non-intrusively. This Chapter demonstrates an application of the video quality prediction models developed in Chapter 5 in (1) optimization of content and network provisioning over WLAN and (2) QoS control over UMTS and WLAN networks.

First, a QoS-driven adaptation scheme for optimizing content provisioning and network planning for video applications over wireless networks is proposed in this chapter. Two major research questions for video bitstream adaptation subject to low-bitrate constraints transmitted over error prone and bandwidth restricted wireless access network environment have been looked at:

(1) How content providers can provide the optimized video content matching user's QoS requirement?

(2) How network providers can best utilize existing network resources according to user's QoS requirement?

The first question is addressed by adapting the SBR according to a video quality prediction model described in Chapter 5. This ensures a maximization of users QoS. The second question is addressed by finding the impact of encoder related parameters of SBR and Frame Frame Rate (FR) and WLAN access network parameters of Packet Error Rate (PER) on delivered QoS in order to maintain acceptable quality.

Second, an application of the models in QoS control is given by adapting the sender bitrate according to user's QoS requirements.

The structure of the Chapter is as follows. Section 7.2 presents an overview of relatedwork to sender bitrate adaptation schemes for the optimization of users' perception of service quality. In section 7.3, the proposed QoS-driven scheme for optimization of content provisioning and network resources is introduced which is based on the simulation set-up given in Chapter 5, sub-section 5.3.1 In section 7.4, the QoS-driven fuzzy adaptation scheme is introduced over UMTS and WLAN. Section 7.5 presents the results and analysis of the QoS-driven fuzzy scheme. Section 7.6 summarizes this chapter.

7.2 Related work

The provision of optimized QoS is crucial for wireless/mobile multimedia design and delivery. With limited resources and bandwidth constraints video adaptation have become one of the most important and challenging issues in wireless/mobile multimedia applications. Perceived QoS is crucial in the service uptake by users and hence in the full utilization of the potential services that are offered as a result of advancement access network and multimedia technologies. Users' demand for quality of video applications is very much content dependent and streaming video quality for example, is dependent on the intrinsic attribute of the content. Several researchers have proposed adaptation schemes in literature. Work presented in [132] proposed a content-based video adaptation scheme where content features are extracted using a machine learning method from compressed video streams. The content features are then used in the adaptation operation. In [94] a content-based adaptation scheme using an optimum adaptation trajectory was proposed. In [133],[134] video adaptation based on utility function obtained from content characteristics was proposed. They propose a model based on utility functions in terms of the send bitrate and frame rate for each video sequence.

In [135] a context-aware computing to adapt video content accessed by users with differing device capabilities has been presented. Work presented in [136] insists that the use of contextual information is essential to achieve efficient adaptations that can enrich the user experience. They present a scalable and modular platform for context-aware adaptation of multimedia content that is governed by digital rights management. In [137] the problem of optimizing the delivery of multimedia services is given by assuming that the network provides two distinct classes of service to users based on their QoS requirement such as premium or economy. Based on that they presented a filtering strategy that adaptively controlled the allocated bandwidth and the transfer delay of traffic flows downloaded from content servers to mobile users. In [138],[139] a bitrate control scheme based on congestion feedback over the Internet has been proposed. In [140],[141],[142] adaptation based on network state and congestion control over UMTS transport channels has been presented. Work in [143] have presented an adaptive bandwidth allocation scheme based on the queue length and the packet loss probability. In [144],[145] a rate allocation algorithm for H.264 video has been proposed that exploits the Gilbert-Elliot model at macroblock level. Work presented in [146] discusses the effects of sender bitrate adaptation using asymmetric dropping of quality enhancement layers from key frames and non-key frames exploring H.264/SVC. In [147] the end-to-end QoS provisioning for scalable video streaming over a heterogeneous IP/UMTS networks has been discussed. In [148],[149] an adaptive fuzzy rate control feedback algorithm based on packet loss rate and congestion notification from routers has been presented. They combined network adaptation techniques with content adaptation techniques to achieve performance degradation when the network load increases. In [150] a model is proposed based on dynamic bitrate control to subjectively estimate the quality of video streaming. Their estimation model considers user perception in three areas where quality degradation is high, the impression of past quality and the duration of degradation. A scheme based on packet dispersion instead of packet loss is presented in [151] using a fuzzy rule in combination with a transcoder to adapt the video bitrate. In [152] concepts behind framework supported by MPEG-21 (part 7) to enable terminal and network QoS has been presented. Similarly, in [153] based on the MPEG-21 framework adaptation from their model based on RTP traffic data has been proposed. Whereas, [154] presents a Multi-user Session Control (MUSC) to provide QoS mapping and QoS adaptation of multi-user sessions over heterogeneous and mobile networks. Work in [155] presents an approach for mapping user experience in terms of MOS to network conditions on their developed next generation network test bed. Their main focus is user roaming scenario across diverse wireless technologies. Video adaptation to suitable three dimensions based on spatial and temporal feature combination is proposed in [15] for scalable video. They adapt the video resolution (from CIF to QCIF) and frame rate based on their proposed quality prediction model. The model had around 74% accuracy. A QoE management method is proposed in [156], where network resources are preserved in a way that minimizes the impact on the QoE. They show a practical demonstration on how to control application QoS parameters for QoE management. The concept of adaptation from perceptual model is given in [157] where a Quality-Oriented Adaptation Scheme (QOAS) for delivering multimedia streams is proposed. Their adaptive mechanism uses feedback from clients regarding the quality of delivery to assist the server in making dynamic adjustments to the transmitted streams. In [158] a Region of Interest-based Adaptive scheme (ROIAS) is presented which adjust differently the regions within each frame of the streamed multimedia content based on the user interest in them. Work presented in [159] propose a new perspective into video content adaptation by considering a triple emotion-perception-emotion user model. The model allows widening the set of adaptation solutions for multimedia services targeting the highest user satisfaction. In [160] a semi-fuzzy rate control algorithm for variable bitrate video applications has been presented. They implemented their algorithm in H.264 coded videos. Their semi-fuzzy algorithm uses a quantization parameter to compute the bitrate budget.

The optimization of perceived QoS is crucial for mobile multimedia design and delivery. Most of these schemes maximize network resources without considering the impact of content. Traditional QoS adaptation schemes do not take into account the video content even though video content dynamics is critical for the final perceptual outcome. In addition, the main aim of most of these schemes is to minimize the end-to-end packet loss/delay. This optimization is based mainly on Network Quality of Service parameters without taking the Application QoS parameters in to account.

In existing literature of video quality adaptation there is limited work on adapting the quality from perceptual model. The models proposed in these works for adaptation are trained from raw video features. There is very little work that aims to combine parameters in both layers considering for all content types.

7.3 QoS-driven optimization of content provisioning and network resources

This section describes the proposed adaptation scheme to optimize contents provisioning and network resources.

7.3.1 Introduction to the scheme

The optimization of the content is carried out by applying the QoS-driven model described in Chapter 5 at the receiver end in real time. The optimization of the network resources is dependent on finding the impact of QoS parameters on end-to-end quality for each type of video application. Through statistical analysis of ANOVA and PCA, the QoS parameters have been identified for each video application, hence enabling in the optimization of existing network resources. This is best explained by the flow diagram of the proposed QoS-driven adaptation scheme which is depicted in Fig. 7.1. From Fig. 7.1 the video application is first defined based on the content features (e.g. SM, GW or RM). Then from the video quality prediction model the process of optimization of content provisioning and network resources takes place based on either adapting the SBR or finding the impact of QoS parameters.

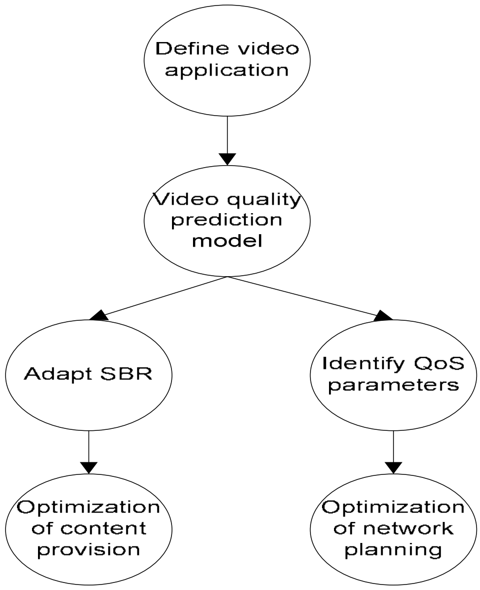


Figure 7.1. Flow diagram of the proposed QoS-driven scheme for optimization of content provisioning and network resources

QoS-driven prediction model

In order to establish the initial encoding sender bitrate for the three content types, the variables of FR and PER are fixed in Eq. (5.19). The FR was fixed at 10fps and PER as 0 assuming that there are no network losses. The SBR versus MOS curve is shown in Fig. 7.2 for the three content types (Fig. 5.6 is re-drawn to reflect SBR vs MOS as opposed to PSNR). The purpose of Fig. 7.2 is to show the maximum and minimum SBR achieve able highlighting the initial encoding requirement. Therefore, it shows the relationship of MOS with application QoS parameter of SBR. From Fig. 7.2 it is observed that there is a minimum sender bitrate for acceptable quality (MOS>3.5) for all content types. A MOS of 4 is considered "good" for streaming applications [161] where most users are satisfied. There is also a maximum sender bitrate for the three content types that gives maximum quality (MOS ~ 4.2). For example for the content category of SM, sender bitrate of 100kbps gives a maximum of 4.2. However, in RM higher sender bitrates are required for maximum quality i.e. > 500kb/s. From Fig. 7.2 it can be derived that when the sender bitrate drops below a certain threshold that depends on the type of video content, then the quality becomes 'bad'. Moreover, the quality does not improve considerably for sender bitrates higher than a specific threshold, which also depends upon the spatial and temporal activity of the video clip. This is useful when applying adaptation to SBR as it defines the initial encoding bitrates for all content types.

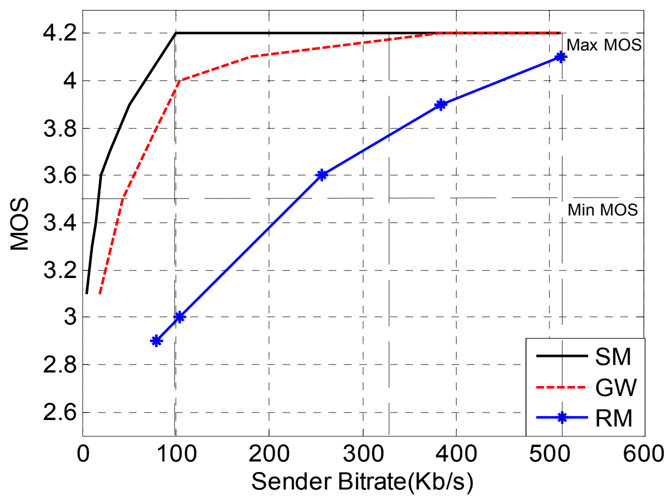


Figure 7.2. MOS Vs Sender Bitrate for the three content types

The reference-free video quality models over wireless network are described in Chapter 5, Section 5.7 where video quality in terms of the MOS is predicted from a combination of encoder related parameters of Sender Bitrate (SBR), Frame Rate (FR) and WLAN access network related parameters of Packet Error Rate (PER) for three different video applications classified earlier as SM for video conferencing application, GW representing a typical video call and RM representative of video streaming (CT=0.5 in Eq. 5.19). The video codec for these applications was MPEG4. The prediction model is obtained by nonlinear regression analysis of the QoS parameters both in the application and network level and is given as below in Eq. (5.19).

$$MOS = \frac{\alpha + \beta FR + \gamma \ln(SBR)}{1 + \mu(PER) + \sigma(PER)^2} \qquad (5.19)$$

For the purpose of demonstrating adaptation, no packet losses are assumed and hence PER = 0. The above Equation then reduced to Equation 7.1 below.

$$MOS = \alpha + \beta FR + \gamma \ln(SBR) \qquad (7.1)$$

The metric coefficients were re-fitted by non-linear regression of the prediction model with our training set (MOS values). The re-fitted metric coefficients α, β and γ along with the correlation coefficient showing the goodness of fit and RMSE for all three video applications over WLAN networks are given in Table 7.1.

Table 7.1 Re-fitted Metric coefficients

Coefficients	SM	GW	RM
α	2.797	2.273	-0.0228
β	-0.0065	-0.0022	-0.0065
γ	0.2498	0.3322	0.6582
R^2	88.82%	89.19%	99.57%
RMSE	0.1399	0.1354	0.0352

The model was trained with three video sequences of Akiyo, Foreman and Stefan in the three categories of SM, GW and RM, whereas the model is verified with three different video sequences of Suzie, Carphone and Football in the three corresponding content categories. MATLAB™ function *nlintool* has been used to carry out the nonlinear regression analysis. R^2 indicates the goodness of fit of the fitted coefficients of the three models. See Chapter 5 for details.

The predicted video quality metrics are then used in the QoS-driven adaptation scheme to adapt the video sender bitrate as shown in Fig. 7.3.

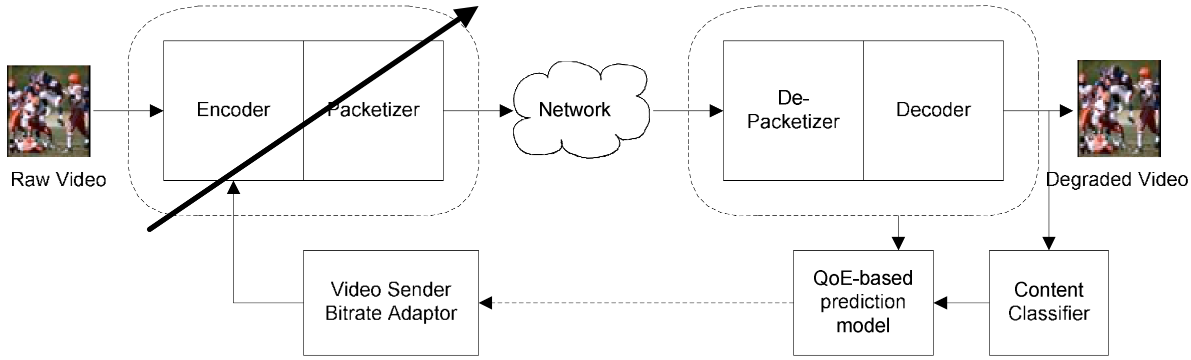


Figure 7.3. Block diagram of the proposed scheme

The basic model of our adaptation scheme, given in Fig. 7.3 consists of the following modules.

<u>Content classifier:</u> The content classifier classifies video sequences into three categories as SM, GW and RM using cluster analysis based on the spatial and temporal features of the video and hence determines the content type. . The category of SM represents video with low spatial and temporal movement, whereas, the category of RM represents video with high spatial and temporal movement. The third category of GW represents video of low/high spatial and high/low temporal movement. This is described in detail in Chapter 4.

<u>Video Sender Bitrate Adaptor</u>: This block adjusts the sender bitrate of the transmitted video as per the information received from the video quality prediction model given by Eq. (7.1) and according to Table 7.2.

<u>Encoder</u>: Lastly the adapted video clip is achieved according to the quality based on video content dynamics. Layered encoding is used for adapting the video streams to the content dynamics. The video clips are encoded in number of layers in a way that each layer increases the video quality of the video clip. Base layers are encoded at a very low rate to accommodate for different access networks (e.g. UMTS or WLAN). Additional layers are added to adapt the video stream according to the content type.

Table 7.2 Experimental video scale assignment for SBR

Scale Value	Average Sender Bitrate (Kb/s)
1	1 – 43
2	44 – 79
3	80 – 127
4	128 – 255
5	256 – 339
6	≥340

The histogram of objective preference count (MOS values obtained from PSNR conversion [21] for all three content types is shown in Fig. 7.4. Fig. 7.4 shows that each content type has its unique patterns of adaptation preference supporting our adaptation scheme. The video clips are labelled as SM, GW and RM with low, medium and high complexity, respectively. A clear trend can be seen as the video content complexity increases outlining the adaptation preference as more complex videos need higher send bitrate for acceptable quality.

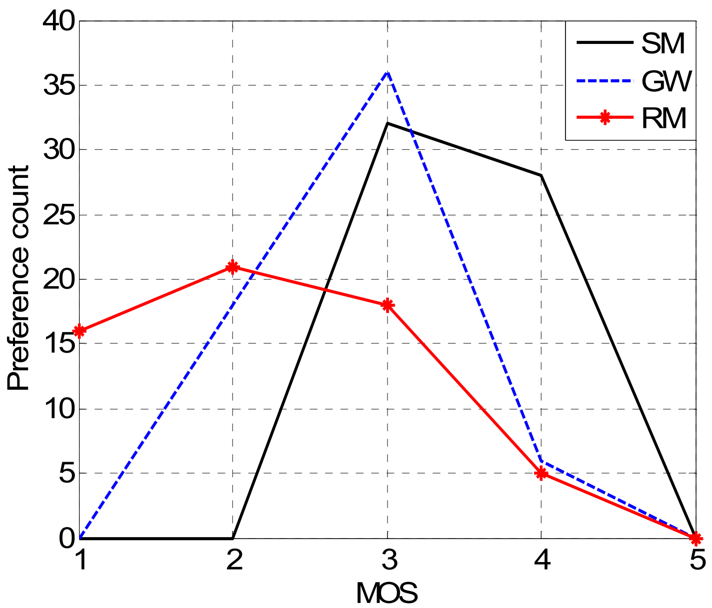


Figure 7.4. Histogram of the three content types

Sub-sections 7.3.2 show the application in content optimization and 7.3.3 show the application in network optimization.

7.3.2 Optimization of content provision

The MOS value is specified by the content provider to achieve specific quality level to meet the end customers' requirement. In today's network infrastructure the motivation for service providers' to provide new services to customers is reduced due to low revenue margins. Therefore, it makes sense to optimize existing network infrastructure and provide service differentiation to customer in terms of premium and tailor-made services according to customer's requirement. Hence, the motivation to use MOS as an input in our model as QoS is best captured in the MOS value. The video application is either video-conferencing, video call or video streaming. The Frame Rate (FR) is decided based on the service and application provider. For demonstration purposes, the application is wireless/mobile environment and hence the FR was fixed at 10f/s. The PER information can be provided from the network statistics and it is assumed no packet loss conditions for concept proofing and simplicity.

The model given by Equation 7.1 is used to obtain the SBR as shown in Fig. 7.5. Therefore, Equation 7.1 can be re-written as Equations 7.2, 7.3 and 7.4 for the three content types respectively.

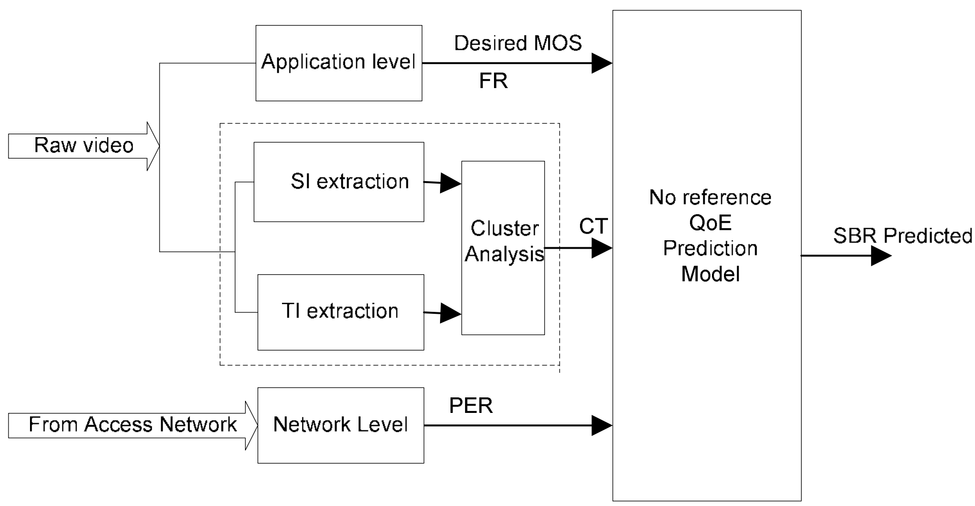


Figure 7.5. Method to calculate the SBR

$SBR_{SM}= e^{(MOS- 2.797+ 0.0065FR)/0.2498}$ *(7.2)*

$SBR_{GW}= e^{(MOS- 2.273+0.0022FR)/0.3322}$ *(7.3)*

$SBR_{RM}= e^{(MOS+0.0228+0.0065FR)/0.6582}$ *(7.4)*

According to [161] for video applications MOS between 3.7- 4.2 is considered ‘acceptable to good’ for most people. From the basic MOS requirement of 3.5, and from Equations (7.2)-(7.4) SBR adaptation is illustrated from content providers point of view in Table 7.3 for the three content types for a quality range of 3.5 - 4.2. Hence the content provider is able to identify the SBR that corresponds to a given QoS (in terms of MOS) level by simply using Equations (7.2)-(7.4) (i.e. video quality -prediction model).

Table 7.3 Predicted Send Bitrate Values for Specific Quality Levels

MOS	FR	SBR_{SM}	SBR_{GW}	SBR_{RM}
3.5	10	21	43	233
3.7	10	48	78	315
3.9	10	107	143	428
4.1	10	237	261	578
4.2	10	354	353	675

7.3.3 Optimization of network planning

Optimization of network provisioning is carried out by finding the impact of the QoS parameters on end-to-end quality. The experimental set-up is the same Section 4.6.1 (Chapter 4). The impact of QoS parameters of SBR, FR and PER is found by carrying out PCA and ANOVA analysis as described below.

Principal Component Analysis

To find the impact of each QoS parameter statistical analysis of Principal Component Analysis (PCA) [83] was carried out. The depth (called dimensions) of the data set is reduced by PCA, such that most of the variables are related to each other. PCA uses a covariance matrix in the case where the same data has the same set of variables or correlation matrix in the case where data has a different set of variables. In this thesis, a covariance matrix was used because of the same data set. The main aim of principal component analysis is to reduce the number of variables of the dataset by retaining most of the original variableness in the data. Most of the variation in the data is accounted by the first principal component followed by each succeeding component that then accounts for the remaining variability. Fig. 7.6 shows the variance of the six principal components for the data. 83% of the variability is explained by the first two principal components of which a total variation by first principal component was 72.7% and 10% by the second component. Consequently, only scores from the first component were chosen.

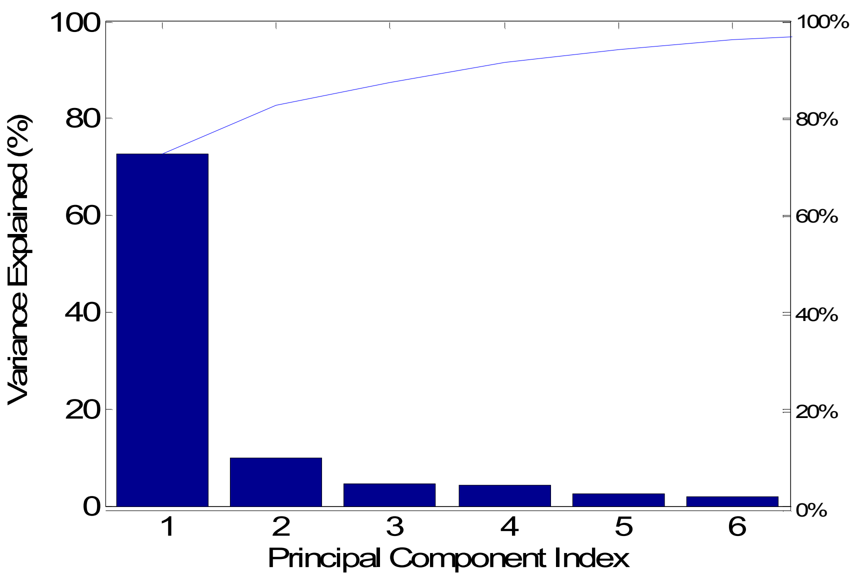


Figure 7.6. Eigen values of the six principal components

The principal component scores for each content is shown in Table 7.4. Table 7.4 shows the influence of each QoS parameter on video quality. The PCA scores of each QoS parameter in Table 7.4 is given under the columns of SBR, FR and PER. The higher the principal component score value (e.g. for the video sequence of Akiyo in the category of SM principal component score of SBR=0.57, whereas, for FR and PER the principal component scores were -0.58) shows that parameter has a higher impact. This shows that for Akiyo the parameter of SBR has a greater impact on quality compared to that of PER and FR as the value of SBR is the highest. Similarly, for the video sequence of carphone, the impact of PER is slightly higher than SBR, whereas FR is least important. To summarize, scores for sports video contents are higher than those of news type videos. Also in the category of RM higher packet loss have a greater impact on video quality compared to that of SBR and FR. Similarly, for SM content type PER does not have a bigger impact on video quality.

Table 7.4 Principal Component Score Table

Content type	Content	Scores	SBR	FR	PER
SM	Akiyo	0.212	0.57	-0.58	-0.58
	Suzie	0.313	0.66	0.25	-0.71
	Grandma	0.147	-0.76	0.64	-0.05
	Bridge-close	0.092	0.41	-0.22	-0.89
GW	Table Tennis	0.287	0.08	-0.99	0.11
	Carphone	0.154	0.35	-0.93	0.10
	Foreman	0.204	0.56	0.45	-0.69
	Rugby	0.454	0.65	-0.59	0.48
RM	Tempete	0.231	0.25	-0.46	-0.85
	Coastguard	0.221	0.62	-0.60	0.51
	Stefan	0.413	0.40	-0.72	0.58
	Football	0.448	0.62	-0.57	0.55

From Table 7.4 the main QoS parameters that impact on end-to-end quality for the three content types are summarized below:

- The main factors degrading objective SM video quality are frame rate and sender bitrate. However, for the sequence of Grandma SBR is a bigger degrading factor compared to frame rate. However, for most sequences in this category the requirements of frame rate are higher than of sender bitrate.
- The main factors degrading objective GW video quality are the sender bitrate and packet error rate. In this category packet loss has a much higher impact on quality compared to SM.
- The main factor degrading the RM video quality are sender bitrate and packet error rate. A video coded at low sender bitrate and/ with high packet losses is very

annoying for most users. This is because the initial encoding requirement of fast moving video is greater than slow moving video. This is shown in Fig. 5.2 previously, where for content type of RM minimum MOS is achieved for a sender bitrate of 220kbps compared to that of SM of around 30kbps. Also if packet losses are high, that results in partial/total loss of I-frames thus reducing the overall end-to-end quality.

ANOVA Analysis

In order to thoroughly study the impact of the QoS parameters on MOS, ANOVA (analysis of variance) [83] was performed in addition to PCA on the MOS data set. A three-way repeated measurement analysis of variance (ANOVA) on the MOS data set given by the 3 QoS parameters differ when grouped by multiple factors (i.e. the impact of all the factors combined) is given in Table 7.5 (Table 3.6 has been extended to account for 2-way interactions) and supports the observations concerning the impact of the QoS parameters of SBR, FR and PER. Table 7.5 shows the results, where the first column is the Sum of Squares, the second column is the Degrees of Freedom associated with the model, the third column shows the Mean of Squares, the fourth column shows the F statistic and the fifth column gives the p-value, which is derived from the cumulative distribution function (cdf) of F. The small p-values ($p \leq 0.01$) indicate that the MOS is substantially affected by at least one parameter. The results of ANOVA reported in Table 7.5 indicate the main effects are due to SBR and PER (p-value=0). The FR on the other hand, is not as significant as SBR and PER. There were interactions between each pair of factors. Specifically, the p-value ($p=0.4573$) for the two way interaction between frame rate and sender bitrate indicates that the impact of frame rate with SBR is not as significant as the impact of SBR itself. Hence, it is important to achieve an optimal SBR-FR trade-off for acceptable quality. The two way interaction of SBR and PER gave 0, which shows that those are two most important QoS parameters.

Table 7.5 ANOVA Results for Main and Interaction Effects

Source	Sum of squares	df	Mean Squares	F-value	p-value
Send Bitrate (SBR)	108.966	4	27.2414	82.91	0
Frame Rate (FR)	1.681	2	0.8406	2.65	0.0718
Packet Error Rate (PER)	105.226	4	26.3066	82.96	0
SBR * FR	2.466	8	0.3082	0.97	0.4573
FR * PER	3.386	8	0.4232	1.33	0.2244
SBR * PER	24.26	16	1.5163	4.78	0

Results showing optimization of network resources using two content types

Two cases were demonstrated to show how the proposed QoS-driven adaptation scheme benefits both the provisioning of existing network resources.

The video sequence of 'Suzie' from the content type of SM is encoded at 10f/s with sender bitrate of 40kb/s. It is known from Table 7.4 that the sender bitrate can be reduced while maintaining the same quality. The sender bitrate was reduced to 20kb/s. The network throughput utilization is depicted in Fig. 7.7. Approximately 16% gains of network resources are achieved. Similarly, Fig. 7.8 shows a very small gain of 6-8% for the content type of RM by increasing the SBR from 384kb/s to 512kb/s and reducing the frame rate from 30f/s to 10f/s and hence the optimum trade-off to maintain QoS. The increase in SBR actually increases the network bandwidth utilization. However, this is compensated by reducing the frame rate. Therefore, as a result there is negligible net gain on network resources in this case compared to the previous one shown in Fig. 7.7.

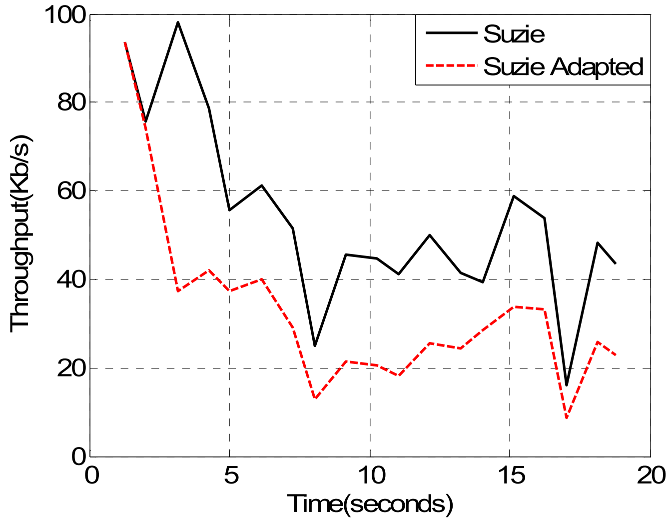


Figure 7.7. Network throughput utilization for SM

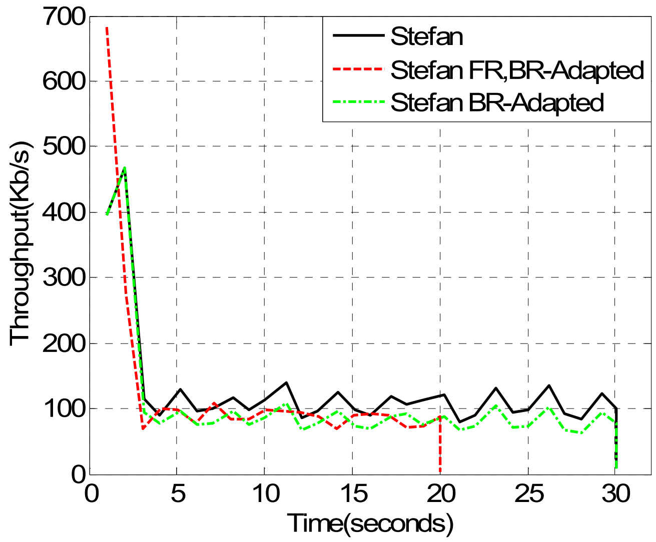


Figure 7.8. Network throughput utilization for RM

It has been shown from Figs. 7.7 and 7.8 that user's QoS can be maximized while preserving network resources. The results from Table 7.4 allows network operators to allocate network resources according to user's QoS requirements.

7.3.4 Comparison to existing work

Recent studies in [15] have proposed video adaptation to suitable three dimension combination based on spatial and temporal feature combination. Whereas, work in [156] proposed a QoE management methodology aimed at maximizing user's QoE and in [63] authors propose an optimization function that adapts the bitrate of each application, e.g. voice or video, data, etc. that share the same QoS requirement and hence optimizes the network throughput for the delivery of multimedia services with different QoS requirement. Scheme in [15] focuses on the application level parameters only for adaptation. In [63] authors main focus is adapt the network bandwidth according to requirement of the user, e.g. premium or economy. Scheme in [63] focuses mainly on the network level. Whereas in [156] authors propose a QoE-aware management system that maximizes existing network resources. Authors in [156] have demonstrated how AQoS can be used to maximize network resources. Compared to these schemes, the proposed scheme addresses the optimization in both application and network level by optimizing both content provisioning and network resources. The scheme is driven by user's QoS by the proposed prediction model that uses a combination of application and network level parameters. In the application level the proposed scheme demonstrates the application of the model in maximizing content provisioning by adapting the SBR. In the network level the proposed scheme demonstrates the utilization of existing network resources by finding the impact of QoS parameters (both AQoS and NQoS) and finding an optimal trade-off between them. The proposed scheme enables a content provider to estimate the delivered video considering specific encoding parameters and network requirements. In addition, it also enables the network provider depending on the impact of QoS parameters to allocate network resources intelligently.

7.4 QoS-driven SBR adaptation scheme over WLAN and UMTS

This section describes the QoS-driven adaptation scheme. The results and analysis over WLAN and UMTS are presented in the next section (7.5).

7.4.1 Introduction to the scheme

In this section the application of the proposed model in sender bitrate adaptation at the sender side is described. The advantage of the fuzzy logic [162] was taken and applied to the adaptation scheme. Fuzzy controllers are generally used in situations where it is difficult to obtain formal analytical models as rigorous control theoretical approaches cannot be applied. There are mainly four modules in a fuzzy logic controller. This is shown in Fig. 7.9. They are called fuzzification module, defuzzification module, fuzzy inference engine and fuzzy rule base. The most important ones are the linguistic variables and the membership functions. It is the linguistic variables that explain the attributes of the variables using if-then linguistic terms. These linguistic terms are then characterised by a fuzzy set which is the membership function. A membership function is a curve that explains how each linguistic term is linked to a membership value (or degree of membership) between 0 and 1. The most common membership functions are triangular ones (trimf) chosen because of their simplicity.

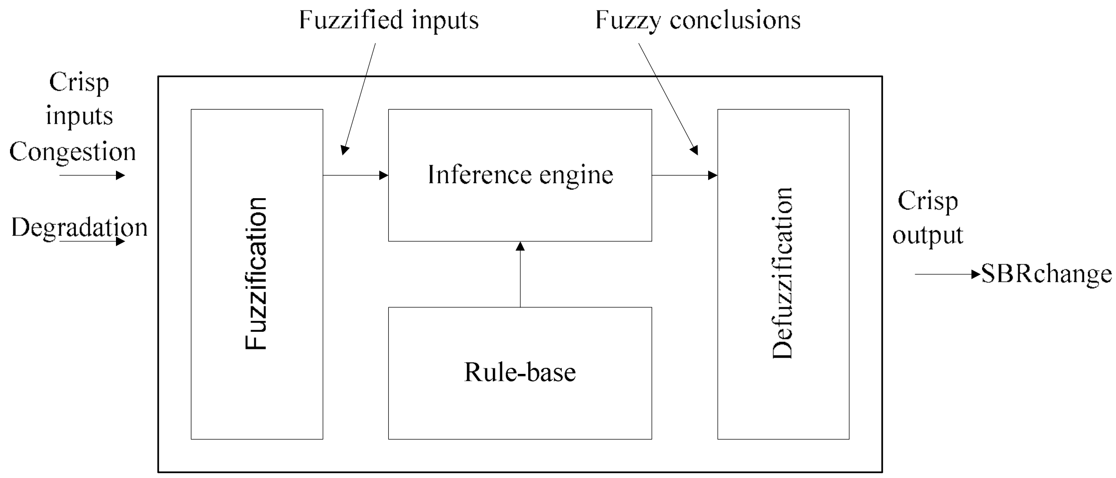


Figure 7.9. Fuzzy sender bitrate adaptor concept

In Fig. 7.9, the fuzzification module takes the crisp inputs and converts them into fuzzified inputs. These fuzzified inputs are then taken by the inference engine, where the if-then linguistic rules are applied to them. The output of the inference engine are fuzzy conclusions which are taken by the defuzzication unit and converted back into crisp output.

Fig. 7.10 illustrates how the video quality is predicted non-intrusively and shows the concept of QoS-driven adaptation. At the top of Fig. 7.10, intrusive video quality measurement block is used to measure video quality at different network QoS conditions (e.g. different packet loss, jitter and delay) or different application QoS settings (e.g. different codec type, content type, sender bitrate, frame rate, resolution). The measurement is based on comparing the reference and the degraded video signals. Both objective measurements from PSNR and Subjective MOS values are used for measuring video quality in this thesis (See Chapter 5). The video quality measurements based on MOS values are used to derive non-intrusive video quality prediction model and sender bitrate adaptive control mechanism based on non-linear regression methods as described in Chapter 5.

Adaptation can be either at the sender side, such as adapting the sender bitrate or the receiver side e.g. from buffer, to achieve an optimized end-to-end QoS. In Fig. 7.10 the video content classification is carried out from degraded video at the receiver side by extracting their spatial and temporal features. The spatio-temporal metrics have quite low complexity and thus can be extracted from videos in real time. Video contents are classified as a continuous value from 0 to 1, with 0 as content with no movement e.g. still pictures and 1 being a very fast moving sports type of content. The content features reflecting the spatio-temporal complexity of the video go through the statistical classification function (cluster analysis) and content type is decided based on the Euclid distance of the data. Therefore, video clips in one cluster have similar content complexity. Hence, the content classifier takes the content features as input observations, while content category as the output. For larger video clips or movies the input will be segment by segment analysis of the content features extracted. The CT estimation is carried out from the ANSI SI and TI information. The spatial features extracted are edge, blurriness and brightness, whereas, the temporal features extracted are the sum of absolute difference value. Based on the extracted features, cluster analysis was carried out to determine the content type (See Chapter 4). The detail of the content classification design is explained in Chapter 4. Therefore, within one movie clip there will be a combination of all content types.

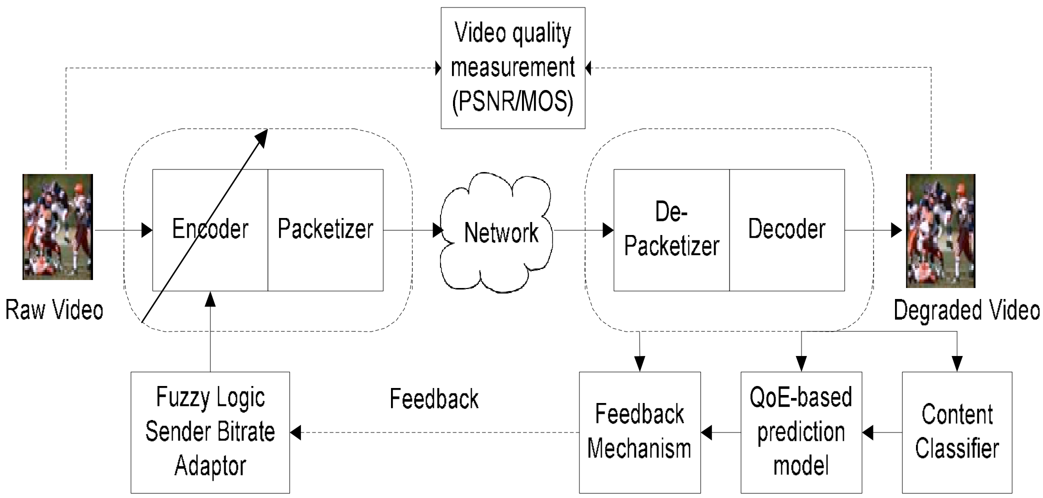


Figure 7.10. Conceptual diagram to illustrate QoS-driven adaptation

7.4.2 QoE prediction model

The video quality prediction model over WLAN and UMTS were derived in Chapter 5 and are given by Equations 5.20 and 5.22 and written below as:

$$MOS = \frac{\alpha + \beta e^{FR} + \gamma CT + \delta \ln(SBR)}{1 + \mu(PER) + \sigma(PER)^2} \qquad (5.20)$$

$$MOS = \frac{\alpha + \gamma ln(SBR) + CT(\delta + \epsilon ln(SBR))}{1 + \mu(BLER)MBL} \qquad (5.22)$$

7.4.3 QoS-driven adaptation scheme

Advantage of the fuzzy logic [162] has been taken which is enforced at the sender side, processes the feedback information and decides the maximum number of layers that will be sent using fuzzy logic control in Fig. 7.10. Layered encoding is used for adapting the video streams to the network dynamics. The encoding of the video streams is done in such a way that each layer increases the perceived quality of the video stream Base layers are encoded at a very low rate to accommodate for the UMTS access network conditions. Additional layers are added or dropped in order to adapt the video stream according to the content type and network conditions.

Two inputs to the adaptation scheme are described in detail as Congestion (C) and Degradation (D).

To calculate the first input, C, the model proposed in Eq. (5.20) over WLAN and Eq. (5.22) over UMTS is used for MOS prediction. The model is light weight and easy to implement. The predicted video quality metrics together with network QoS parameters is then used in the QoS-driven adaptation scheme to adapt the sender bitrate as shown in Fig. 7.10. RTCP is used to exchange the feedback on the quality of the data distribution by exchanging reports between the sender and the receiver. The feedback information is sent through extended RTCP reports [163] every second from the network and collects QoS information like loss rate, delay and jitter from the core network to give the network congestion level. The network congestion level is calculated from the packet loss ratio. Here it is referred to as the Block Error Rate (BLER) computed from the total number of blocks lost over the total blocks sent. BLER is used as opposed to packets lost as in UMTS networks, as it is in the physical layer that transport blocks are passed to the Medium Access Control (MAC) layer. These transport blocks also contain the error indication from Cyclic Redundancy Check. Therefore, the output of the physical layer is then defined by the overall probability of the transport blocks lost or just blocks lost as an error rate (BLER).. Thus, an error model based on 2-state Markov model [54] of block errors was used in the simulation. Further, Congestion (C), is computed from [163] as the fraction of the number of Block/Packets Lost (BL) divided the total number of Blocks/Packets Sent (BS) within an interval. Therefore, the congestion, C, is given by Eq. (7.5) as:

$$C = \frac{BL}{BS} \tag{7.5}$$

The range of congestion level is from [0,1] with 0 being no congestion and 1 meaning fully congested network. The Congestion, C, was partitioned into four levels as ($0<C\leq0.05$), ($0.05<C\leq0.15$), ($0.15<C\leq0.25$) and ($C>0.25$). C is an input to the decision algorithm for SBR adaptation. Each level of congestion corresponds to Blocks lost. $C>0.25$ corresponds to blocks lost of around 7%.

The second input to the decision algorithm is the Degradation (D) and is calculated as the difference between the maximum achievable MOS and the instant MOS_t (computed from the video quality prediction model given in Eq. (5.20) over WLAN and Eq. (5.22) over UMTS). The maximum achievable MOS is set to 4.2 when no blocks/packets are lost. The Degradation, D, is therefore given by Eq. (7.6) as:

$$D = MOS_{max} - MOS_t \qquad (7.6)$$

The maximum value that D (degradation) can have is 3.2 (as the range of MOS is from 1-5), indicating maximum degradation, and the minimum value that D can have is 0 indicating no degradation at all. The degradation, D has been split into four levels as 0-0.25, 0.25-0.7, 0.7-1.2 and $D>1.2$. The split in the values of D are chosen as a change of 0.25 in MOS. This is then linked with an SBR level. The levels of D are chosen such that MOS ranges from 3.8-4.2, 3.8-3.5, 3.5-3.0 and <3.0. The degradation, D, along with the Congestion, C, are used as input to the fuzzy logic sender bitrate adaptor.

The membership functions for the two inputs (linguistic input variables) and the output (SBRchange) is shown in Fig. 7.11. Triangular functions are chosen due to their simplicity. The SBR change (output) surface is also given by Fig. 7.11 which shows the overall behaviour of the SBR adaptor. The first linguistic variable (LV) input C is the network congestion. It ranges from 0 to 1. The second LV, D is the degradation calculated from video quality prediction model. D ranges from 0 to 3.2.

The fuzzy SBR adaptor processes the two linguistic variables based on the predefined if-then rule statements (rule base) shown in Table 7.6, and derives the linguistic output variable SBRchange, which is defined for every possible combination of inputs. An example of the fuzzy rule is:

If congestion is large (L) and degradation is medium (M) then SBRchange is BC (big change)

The linguistic variables in Table 7.6 are given by the membership functions of the output in Fig. 7.11 and are described as No Change (NC), Very Small Change (VSC), Small change (SC) and Big Change (BC). The linguistic variables in Table 7.6 for the two inputs are given by Zero (Z), Small (S), Medium (M) and Large (L). The defuzzified output can then be used to determine the next level of SBR as given by Eq. (7.7).

$$SBR_{new} = SBR_{old} + SBRchange \qquad (7.7)$$

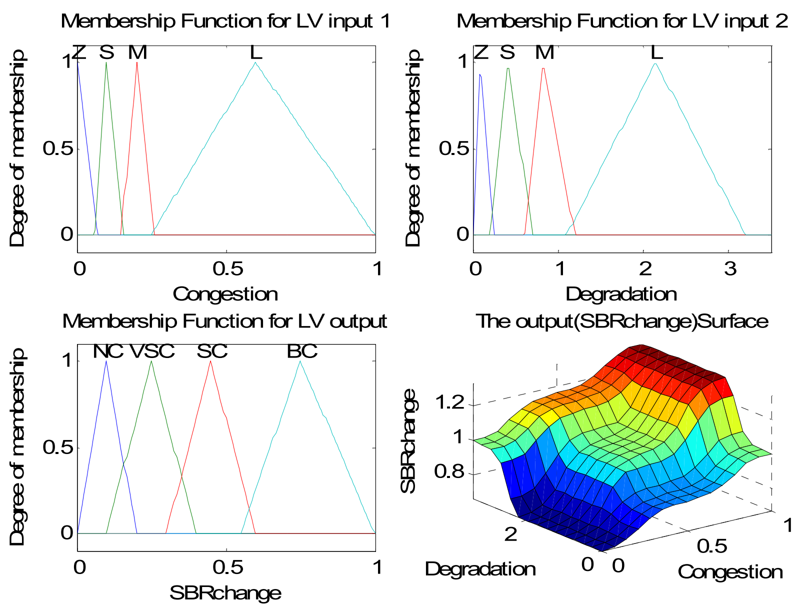


Figure 7.11. Membership functions for the two inputs and the output and the output SBR adaptor surface over UMTS and WLAN

Each value of SBRchange corresponds to a layer of the encoded video bitstream. The defuzzified output is selected from 0 to 1 as shown in Fig. 7.11. Thus a small increase in

SBR is allowed when the bandwidth is available and there is no/reduced congestion, whereas, adaptation takes place to reduce the SBR in case of serious congestion.

Table 7.6 Linguistic Rules

SBRchange		Congestion			
		Z	S	M	L
Degradation	Z	NC	VSC	SC	BC
	S	VSC	VSC	SC	BC
	M	VSC	SC	SC	BC
	L	SC	SC	BC	BC

7.5 Results and analysis of the QoS-driven SBR Adaptation Scheme

This section discusses the results over WLAN and UMTS access networks. The experimental set-up is the same as described in Chapter 3, sub-section 3.3.3 and 3.3.4 for WLAN and sub-section 3.5.3 and 3.5.5 for UMTS. The only difference is that FR is fixed at 30fps.

7.5.1 Results over WLAN

Experiments were conducted with content type of Stefan and assessed the performance of the QoS-driven adaptation scheme over WLAN in terms of MOS. Fig. 7.12 shows adapted vs not adapted stream for video streaming application (content type Stefan) over WLAN and compares to TFRC [164]. With Evalvid-RA framework [165], it is possible to simulate pure TFRC transport directly on top of the network layer. MOS values obtained from PSNR to MOS conversion from [21] are compared to non-adaptive and TFRC. Fig. 7.12 reveals that the QoS-based fuzzy adaptive scheme successfully adapts the sender bitrate according to network congestion. The proposed scheme slowly reduces the sender bitrate according to the network conditions maintaining acceptable quality. TFRC uses a more aggressive manner of recovery after network congestion and increases their transmission rate faster causing significant degradations of end-user perceived quality. There is a clear improvement in quality after adaptation using our proposed scheme.

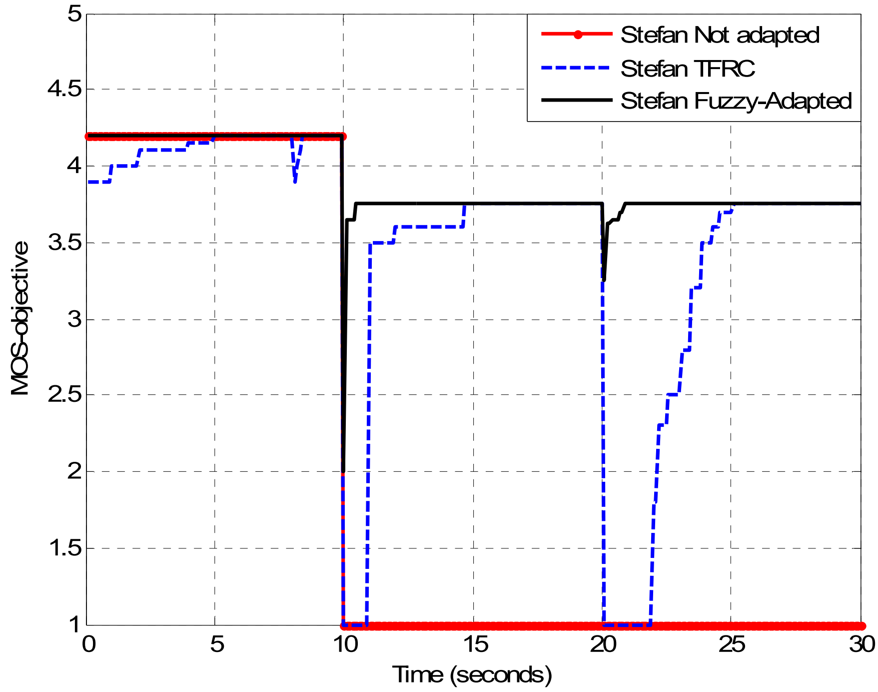


Figure 7.12. Comparison of video quality for adaptive and non-adaptive 'Stefan' video over WLAN

The bottleneck bandwidth was set to different link bandwidths from 50kb/s to 650kb/s, and the performance of the proposed adaptation scheme was appraised and compared to a non-adaptive video.. Fig. 7.13 shows the result of the adaptation scheme. The adaptive scheme does not drop video quality as bandwidth is reduced compared to no-adaptation, adaptive video scheme smoothly adapted the video quality to the network bandwidth that was available over WLAN access network. From Fig. 7.13, it is observed that for a bottleneck bandwidth of 150kb/s, the adaptive scheme gives a MOS of 3.2 compared to 1.2 without adaptation. Fig. 7.14 shows the perceptual quality comparison at 650kbps before adaptation to 80kbps after adaptation for 'Stefan' frame 173. Visually a clear improvement in quality after adaptation can be seen.

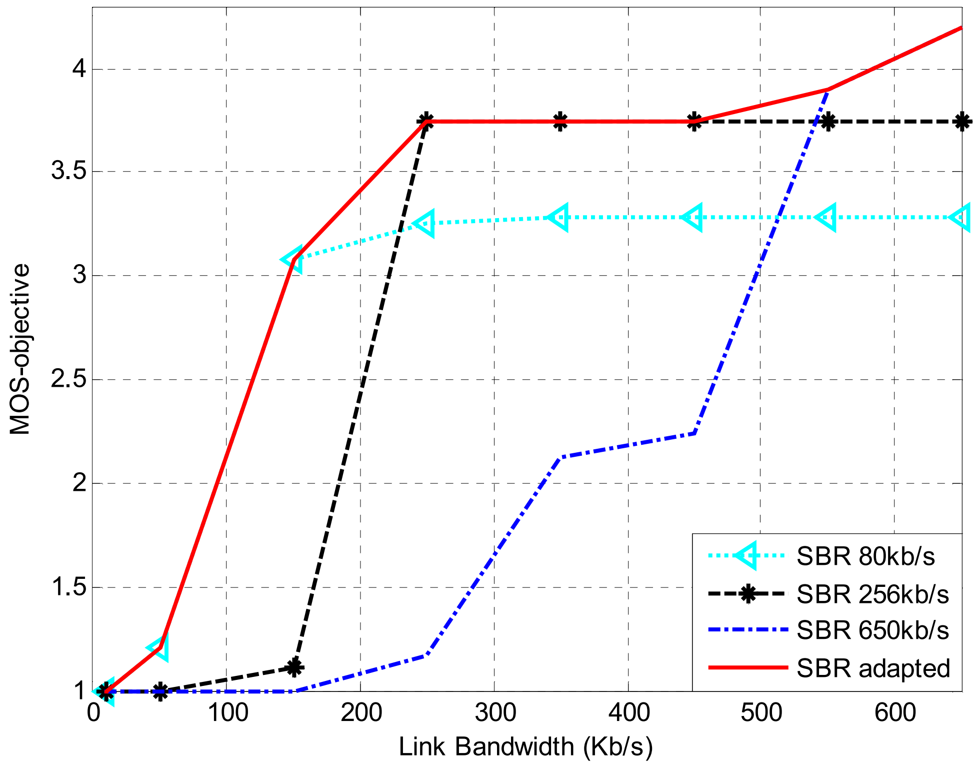


Figure 7.13. Comparison of video quality results for different bottleneck bandwidth

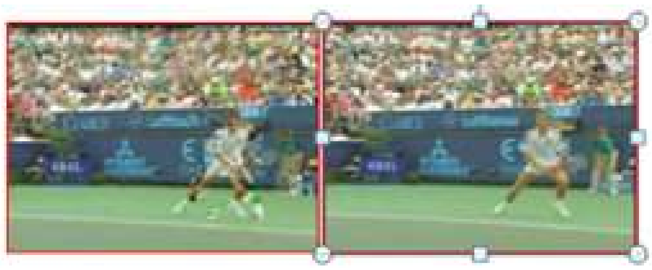


Figure 7.14. Perceptual quality comparison before and after adaptation for Stefan frame 173 10fps, at 650kps to 80kbps after adaptation over WLAN

7.5.2 Results over UMTS

The experimental set-up is the same as described in Chapter 3, sub-section 3.5.3 and 3.5.5. Experiments were conducted with content type of Foreman and assessed the performance of our QoE-driven adaptation scheme over UMTS in terms of MOS. The results of the proposed adaptive scheme are compared with the well known TFRC (TCP-Friendly Rate) [164] controller. The definition of TFRC from [164] is "the sending rate is calculated as a function of the measured packet loss rate during a single round trip time duration measured at the receiver". The sender then calculates the sending rate according to [164]. With the Evalvid-RA [165] framework, it is possible to simulate pure TFRC transport directly on top of the network layer. Constant Bit Rate (CBR) videos are used in the simulation to prove concept. However, the technique can easily be extended to Variable Bit Rate (VBR) videos too.

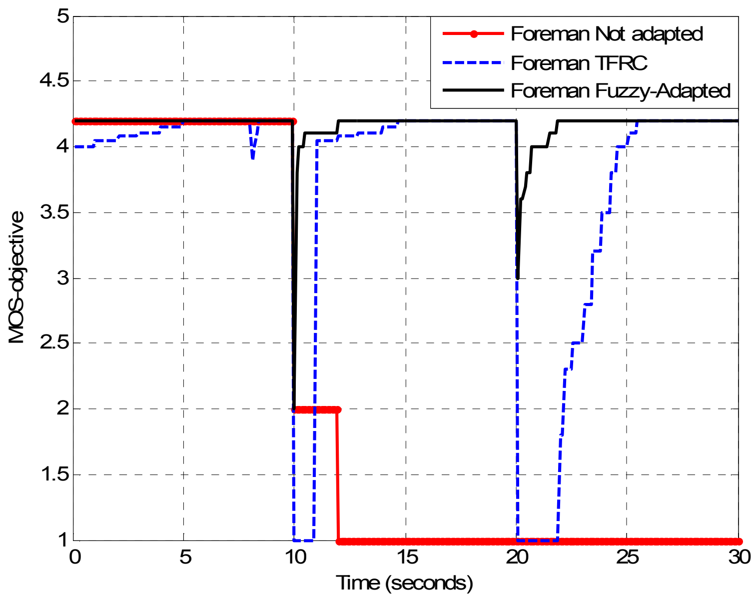


Figure 7.15. Comparison of end user quality with TFRC and no adaptation

The effect of link bandwidth on the MOS (video quality as defined by the user) is analysed by involving one user and then up to five UMTS users were set up that received streaming video over NS2 simulated UMTS network. Experiments were conducted with content type of Foreman and assessed the performance of our QoS-driven adaptation scheme over simulated NS2 [43] UMTS networks in terms of MOS. MOS values are obtained from PSNR to MOS conversion from Evalvid [21]. NS2 was chosen due to its flexibility and based on the characteristics of the link bandwidth. The performance is assessed in terms of user perceived quality (MOS) as shown in Fig. 7.15. MOS values are compared to non-adaptive and TFRC. Fig. 7.15 reveals that the QoS-based fuzzy adaptive scheme successfully adapts the sender bitrate to network congestion. The proposed scheme slowly reduces the sender bitrate according to the network conditions maintaining acceptable quality. TFRC uses a more aggressive manner of recovery after network congestion and increases their transmission rate faster causing significant degradations of end-user perceived quality.

Fig. 7.16 provides a perception for the capacity of the proposed QoS-based fuzzy adaptation scheme with the number of UMTS users that can be sustained by a video streaming server, when the bottleneck link bandwidth is also taken into account. Fig. 7.16 depicts the quality (MOS obtained from PSNR to MOS conversion) that is experienced by up to 5 identical users having the same access network characteristics, especially the bandwidth of the bottleneck link. The dashed lines indicate acceptable quality (MOS~3.5). When the Link Bandwidth (LBW) is high enough to maintain the collective video transmission rate, all users are sustained by the video streaming server at the same video quality levels. Even at the bottleneck LBW of 256kbps all users can be supported at the minimum acceptable level. However, at the lowest level of the LBW only two users can be supported and then the quality reduces below the acceptable threshold.

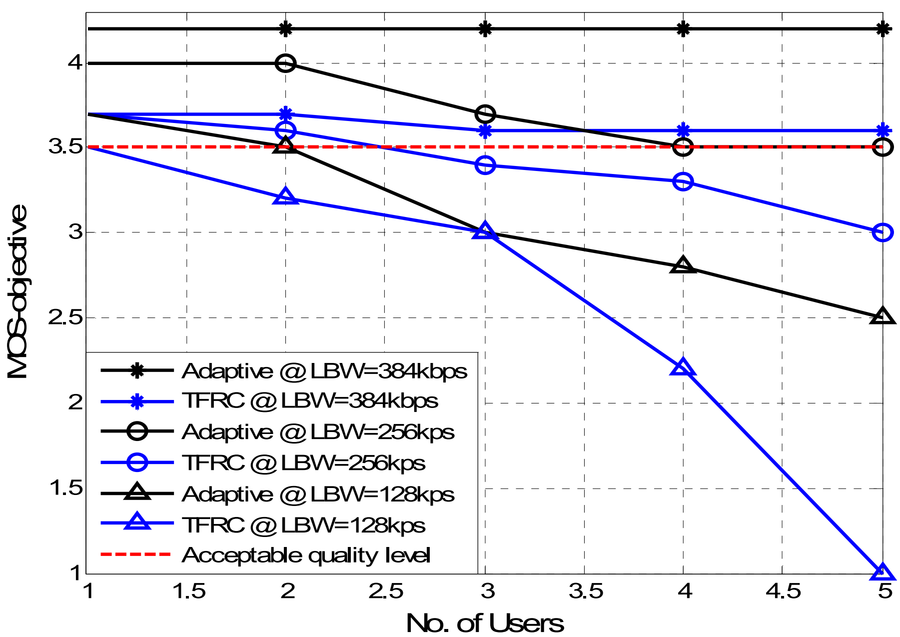


Figure 7.16. MOS-objective vs. Number of active users

Similarly, Fig. 7.17 gives the adaptive video quality over UMTS compared to the non-adaptive one at LBWs of 128kbps, 256kps and 384kbps. Again, an improvement in quality is observed for content type of foreman. At bottleneck bandwidth of 128kb/s, adaptive 'Foreman' gives a MOS of 3.7 compared to 2.7 without adaptation. Therefore, the adaptive video scheme gracefully adapted the delivered video quality to the available network downlink bandwidth.

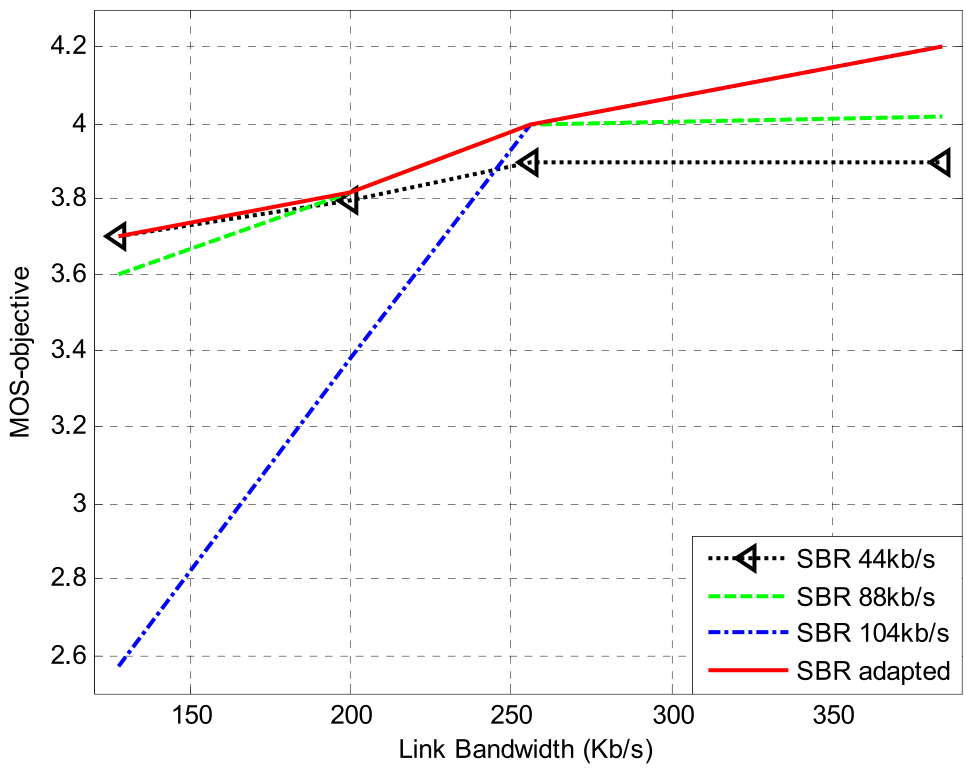


Figure 7.17. Comparison of video quality results for different bottleneck bandwidth over UMTS network

Fig. 7.18 illustrates the advantage of adaptation over UMTS for content type of 'Foreman' for frame 129. An improvement in quality can be seen visually for the frame shown. The SBR is 44kb/s after adaptation and 104kb/s before.

Figure 7.18. Perceptual quality comparison before and after adaptation for Foreman frame 129 10fps, at 104kps to 44kbps after adaptation over UMTS

7.5.3 Comparison of results

From the simulation results of sub-sections 7.5.1 and 7.5.2 the following conclusions can be drawn with regards to adaptation:

1. An accurate video adaptation scheme must consider video content. From the adaptation results, it was found out that the effect of sender bitrate is different for different content types, and hence, it is very important to know the initial sender bitrate to optimize the end user quality.
2. The most important network degradation is packet/block loss. Reducing the SBR does improve quality, if packet loss is high and the reduction in SBR helps to reduce the impact of packet/block loss. The degradations due to reducing the SBR are negligible compared to that of packet/block loss.
3. The access networks have an impact on users' QoS. The effect of packet/block loss over UMTS networks is much higher as it results in long freezes and losing complete GOPs. Also due to the bandwidth restriction, the initial SBR for fast moving content is much lower than WLAN further degrading quality. These results should help in determining network handover strategies in the future.
4. The results show that it is possible to predict video quality if appropriate parameters are chosen. The proposed adaptation scheme responded well to the access network

bandwidth that was available and congestion, outperformed existing TFRC and adapted the SBR accordingly maintaining acceptable quality.

7.6 Summary

This Chapter proposed a QoS adaptation scheme for video applications that maximizes content provisioning and network resources according to user's QoS requirement over wireless local area networks. The video quality -prediction model described in Chapter 5, Section 5.7 has been applied to obtain sender bitrate adaptation and the impact of QoS parameters was found by statistical analysis of PCA and ANOVA. The proposed adaptation scheme enables content providers to identify the video sender bitrates that correspond to various quality levels and hence provide high-quality video services over wireless/mobile networks. It also enables network providers to optimize existing network resources by finding the impact of QoS parameters and hence the trade-off between them.

Further, the prediction model is applied to QoS control. This is achieved by presenting a QoS-driven SBR fuzzy adaptation scheme. The simulation results over WLAN and UMTS are presented and discussed. The proposed scheme is compared to existing TFRC [164] and an improvement is observed in terms of MOS.

Chapter 8

Discussion, Future Work and Conclusions

8.1 Introduction

Transmission of video content over wireless access networks of UMTS and WLAN is growing exponentially and gaining popularity. Mobile video streaming on hand held terminals are seen to be the next killer application and potentially a key factor for their success. However, due to the bandwidth restrictions of UMTS networks video quality still remains of concern. This is because low video quality leads to poor QoS which in turn leads to reduced usage of the application/services and hence reduced revenues.

In order to meet user's QoS requirement, there is a need to predict, monitor and if necessary control video quality. Video quality can be measured by subjective tests or by objective methods. The Mean Opinion Score (MOS) is the most widely used subjective measure of video quality and is recommended by the ITU [1]. The problem with subjective experiments to measure MOS is that it takes a lot of time, are expensive, very difficult to repeat the conditions and cannot be used for real time networks or operations. Hence, objective methods very attractive for real time video quality monitoring in wireless communications access networks. Objective measurement of video quality can be intrusive or non-intrusive. Intrusive methods (e.g. PSNR, SSIM) are widely used, however are less accurate, and are not suitable for checking live video sequences because they need access to the source video. Non-intrusive methods are appropriate for monitoring video quality either directly from wireless access network and/or non-network parameters and hence are preferred. Non-intrusive models provide an effective and practical way to measure user's QoS [101]. The ITU-T standardization sector is actively seeking to standardize video quality prediction models for mobile video streaming applications at the terminal end. VQEG SG9 [20] has recently defined a draft test plan for the hybrid perceptual/bitstream models for mobile video streaming application. The work presented in this thesis is aimed at mobile video streaming applications. This restricts the available bandwidth with current UMTS access networks. Hence, low sender bitrate and resolution (QCIF) was specifically chosen. However, with the advancement in access network technology (e.g. LTE offering up to 20Mbps) and bigger screen resolutions (e.g. smart phones, ipad, etc) now available, the work presented in this thesis reflects worst case scenario.

The main aims of the project are (1) to investigate and evaluate the impact of encoder and access network related parameters for different types of content that affect video quality over wireless access networks of WLAN and UMTS, (2) to investigate methods to classify the video contents objectively, (3) to develop non-intrusive models for the prediction of video quality over wireless access networks of WLAN and UMTS for different types of contents and (4) to apply the developed models in quality optimization and adaptation mechanisms for perceptual QoS control for video over wireless access networks of WLAN and UMTS.

This Chapter discusses the main contributions of this work and highlights the novelty, future work and the conclusions.

8.2 Contribution to Knowledge

The main contributions presented in this thesis are:

(1) Provision of a detailed understanding of the relationships between video quality, wireless access network impairments (e.g. packet/block loss, link bandwidth,

mean burst length), encoder related impairments (e.g. encoder sender bitrate and frame rate) and content type.

The work has contributed to a detailed understanding of the perceptual effects of the key Qos parameters on video quality. This provides a foundation for the development of efficient regression models and artificial neural network learning models. A fundamental investigation to study the impact of the main access network parameters (i.e. packet/block loss rate, link bandwidth and mean burst length), encoder related parameters (e.g. encoder sender bitrate and frame rate) on different types of content on perceived video quality is undertaken using MOS obtained objectively (from PSNR to MOS conversion) and MOS obtained from subjective tests.

The work has been contributed to the research community in the following publications - [22]-[26]. The work is described in Chapter 3.

(2) Classification of video contents using statistical tools

The work has contributed to the investigation of two methods of content classification. In the first method spatio-temporal features are extracted. Based on the ST features content classification using cluster analysis was carried out. This is compared with classifying the video contents from MOS values obtained from objective video quality evaluation. The first method was preferred due to its accuracy.

The work has been contributed to the research community in the following publications - [27], [28]. The work is described in Chapter 4.

(3) New models to predict video quality non-intrusively over wireless access networks of WLAN and UMTS (including non-linear regression and neural network models).

The work has contributed to the development of new models for video quality prediction over wireless access networks of UMTS and WLAN non-intrusively. The prediction was based from a combination of parameters related to the encoder, access network and content types. The models were first developed from MOS values obtained from PSNR to MOS conversion from [21]. Later, subjective testing is carried out and the models are re-trained with the subjective data. Efficient regression-based and neural network based models are developed for wireless access networks of WLAN and UMTS. This avoids time consuming subjective tests. Non-linear regression models are efficient and easy to implement. However, they are static, whereas the neural network models even though more complicated, but can adapt to changing environment of wireless access networks due to their ability to learn.

The work has been contributed to the research community in the following publications - [29][31]-[35]. The regression models are presented in Chapter 5 and the neural network models are described in Chapter 6.

(4) Application of the new non-intrusive video quality prediction models to two areas.

The new non-intrusive models were applied to two important areas to illustrate their usefulness. The first area was in optimisation of content and network provisioning. The scheme enables content providers to select optimal SBR and network providers to dimension network resources such as network bandwidth requirements efficiently. This will ensure an improved user experience, by making the content network-aware and the network content-aware.

The second area was in QoS control. A new fuzzy SBR adaptation scheme that is QoS-driven is proposed. The performance of the scheme is measured in terms of the MOS.

The work has been contributed to the research community in the following publications - [31],[32],[36],[38],[39] The work is described in Chapter 7.

(5) The development of Internet-based subjective MOS tests and availability of MOS results to research community

An internet-based methodology for subjective video quality measurement in terms of the MOS values was developed based on the existing VoIP tests [59] and is presented in Chapter 5. This allows quick assessment of video quality over wireless access networks. Controlled MOS tests were carried out at the UOP and at EHU via the PC and mobile handset. Currently, there are very few subjective video MOS databases available for research community. This is restricting academic research on video quality assessment and prediction. Therefore, the database containing raw/degraded clips, subjective MOS values, encoder and access network based parameters has been made available to the research community at **http://www.tech.plym.ac.uk/spmc/staff/akhan/mos_scores.html** [41].

The work has been contributed to the research community in the following publication [42]. The work is described in Chapter 5.

8.3 Limitations of the current work and discussions

The work carried out in the project has a number of limitations that should be addressed in future studies.

(1) Simulation based performance evaluation

The performance of the proposed algorithms and methods in this project are assessed in simulated networks using mainly the NS2 simulator and OPNET. This approach has benefits of being fast, repeatable, easy to configure and customize. In a simulated network, many parameters such as packet/block loss, bottleneck link bandwidth, etc. are controllable. Simulation based tests are also much more economical than those based on emulation or physical implementation, which involves computing devices and communications interfaces. However, the reliability and consistency of the simulation based tests depends on the quality and accuracy of the simulation models used. In real networks the network conditions are unpredictable.

(2) The accuracy of the subjective tests

Internet based subjective tests were designed and carried out. The two devices were used - PC and mobile handset. The test conditions were designed to take into account the errors associated with the UMTS access networks. However, these tests were limited as real 3G network was not used. Also the PC based tests gave a much bigger screen to subjects compared to a small handheld device. Consequently, the results of this project may not reflect fully the performance of the developed algorithms in practice.

(3) Limited consideration of end-to-end impairments

In this thesis, the impairments considered are mainly access network impairments. However, in reality IP network itself will have impairments e.g. IP packet loss and delay. In addition, in the access network QoS parameters of packet/block loss, mean burst length and link bandwidth were chosen. Delay and jitter were not considered and they may have an impact on end-to-end quality.

(4) Limited consideration of video content types

In this thesis, three types of video contents were considered. They broadly covered video sequences from slow moving (head and shoulder) to fast moving (sports type). However, cartoon clips and movies were not considered. The spatio-temporal features of cartoon and movie clips may have an impact on end-to-end quality.

(5) Limited validation of the work

Although, the models proposed have been validated using an external database, large scale validation or cross validation are still needed. For example, the neural network training set has to be processed according to real data over access networks and to be tested in real time in large scale networks.

(6) Limitation of low resolution and bandwidth

The models proposed in this thesis are aimed at low sender bitrate and resolution videos. With the increase in available bandwidth in the access networks and newer smart phones standardizing a screen resolution of 320x240, the proposed models in this thesis will have to re-trained.

The existing work can easily be applied to higher resolution and bandwidth (e.g. LTE). The models would have to be retrained. The proposed model equation and the ANFIS-based model both would be valid once they have been retrained with the new data.

8.4 Suggestions for Future Work

8.4.1 General directions of future work

There are four main aspects of the research that can be improved and extended further in future work.

1. **Performance evaluation using real system implementations**

The proposed models were implemented in Internet Multimedia Subsystem (IMS) [166] test bed developed at UOP. There have been some preliminary results when the model was applied to the Android G1 Handset [167]. The video quality prediction model as described in Equation 5.22 and adaptation given in Equation 7.7 were embedded in Android G1 [168] and used to compute video quality in terms of MOS which was used for monitoring the video quality over UMTS access network. Due to the lack of UMTS bandwidth control and contractual agreement with 3G Hutchison UK Ltd, no background traffic was introduced but slight video quality degradation were monitored in order to trigger the adaptation mechanism. BLER was periodically retrieved from Android Fieldtest application at an interval of 1 second, and used to predict the quality of the delivered video. Once the MOS begins to drop below 3.2 then adaptation takes place. Fig. 8.1 shows the gain in video quality when adaptation mechanism is in place as opposed to the scenario where there is no adaptation mechanism. The SBR is reduced from 104kbps to 44kbps for Foreman video sequence.

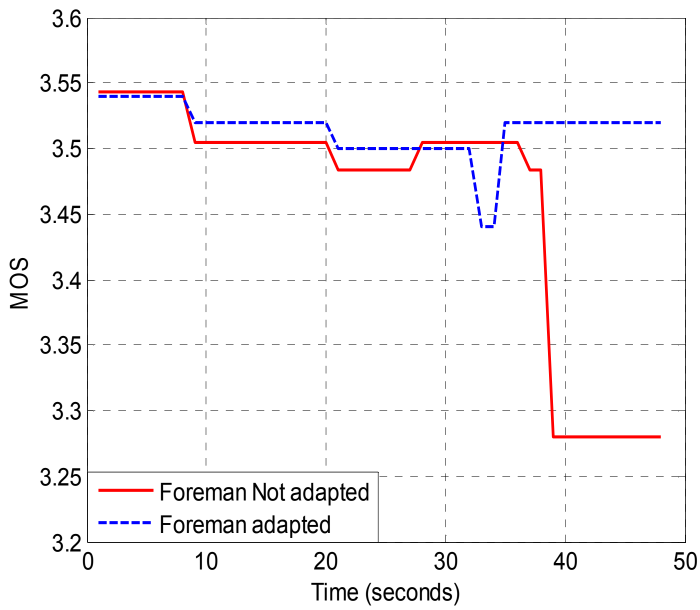


Figure. 8.1. Comparison of 'Foreman' video quality for adaptive and non-adaptive video

Fig. 8.2 illustrates the advantage of adaptation over G1 Android handset for video sequence of 'Foreman'. An improvement in quality can be seen visually for the frames shown.

Figure. 8.2. Perceptual quality comparison before and after adaptation for Foreman over G1 Android test bed

However, more work is required to test the models rigorously. In addition, the simulation based performance evaluation can be addressed in the future by emulation. Emulation brings in some aspects of reality while keeping a certain degree of repeatability, configurability and other advantages of simulation. Emulation can be considered as a compromise solution between simulation and physical implementation in a test bed. However, emulation is still not as popular as simulation in the wireless network research community because of the difficulties in accurately emulating the mobility of mobiles as well as in recreating the desired number of users and the conditions of signal propagations. By using emulations experiments are performed in a semi real environment, i.e. using real operating systems that are operating on real devices and running real applications. The models developed in this project can be tested in an emulation environment to get more convincing results and to observe their performance when being implemented in real devices.

In order to study the behaviour of developed technologies in reality and to validate the simulation/emulation results, their physical implementations in real systems are eventually required as the research matures. The actual performance of the models as well as accuracy of the simulation/emulation experiments can be fully tested as real functioning software. Furthermore, the collected results can be used to prove their usefulness for commercial applications.

2. **Further enhancement of content classification**

The advances of scalable video coding can be applied to the content features extracted to predict adaptation categories e.g. slow moving videos can be automatically selected to adapt to a much lower sender bitrate as compared to fast moving videos. The content classification method 1 proposed in this thesis can be extended to classifying contents continuously from [01] in real time by making use of neural networks to predict contents.

Method 2 proposed in this thesis requires further investigation to find a function that can best fit the content classification using MOS values. The accuracy of the function and its relationship to the spatio-temporal features should also be explored.

3. **Perceived quality of multimedia services (Voice and Video)**

The models developed in the thesis for predicting video quality non-intrusively can be extended to consider voice.

Also the differences between voice and video mean that new parameters that impact on quality have to be taken into account when developing models for predicting audiovisual quality non-intrusively. For example, parameters such as the voice bit rate (BR), etc. have to be considered for voice quality prediction for voice over wireless access networks. Works presented in [169],[170],[171] have discussed metrics for multimedia (audiovisual) quality estimation.

Recent work in [172] discusses the new challenges for subjective assessment of 3DTV – evolution from existing 2D methods. Work done in this thesis could be extended to 3DTV work by considering the additional requirements of 3DTV.

4. **Application of neural network models**

In the thesis, neural network learning models have been presented for predicting video quality, non-intrusively. Neural networks have the ability to learn from the changing network conditions. This property of the neural networks has not been explored in this thesis and would be an area to exploit.

8.4.2 Facilitating multimedia communication over UMTS networks in e-healthcare emergency

Currently £11 billion a year is spent in the healthcare for the elderly. This figure is likely to rise over the next 20 years as the nation is getting older and number of people aged over 65 increases by 40% [173]. The cost of keeping a patient in a hospital bed is over £800 a week, however, if they were cared for at home it would cost around £120 a week. Recently, the UK government has apportioned £100 million for schemes that dismiss bed blocking in NHS trust hospitals. As part of some of these schemes, e-health technology is used that enables patients to recover from operations in their own home rather than in a hospital bed. With the increasing availability of 3G wireless access networks the application in healthcare is growing - be it to provide better access to healthcare professional on the move or in a hospital especially when an emergency has occurred in an area where fixed line communication network is not present.

E-healthcare Application Scenario – Graphical representation

Consider a situation in a roadside emergency/patient being monitored at home/high risk patient wearing wireless sensor that takes place in an environment where fixed computing/communication infrastructure is not available.

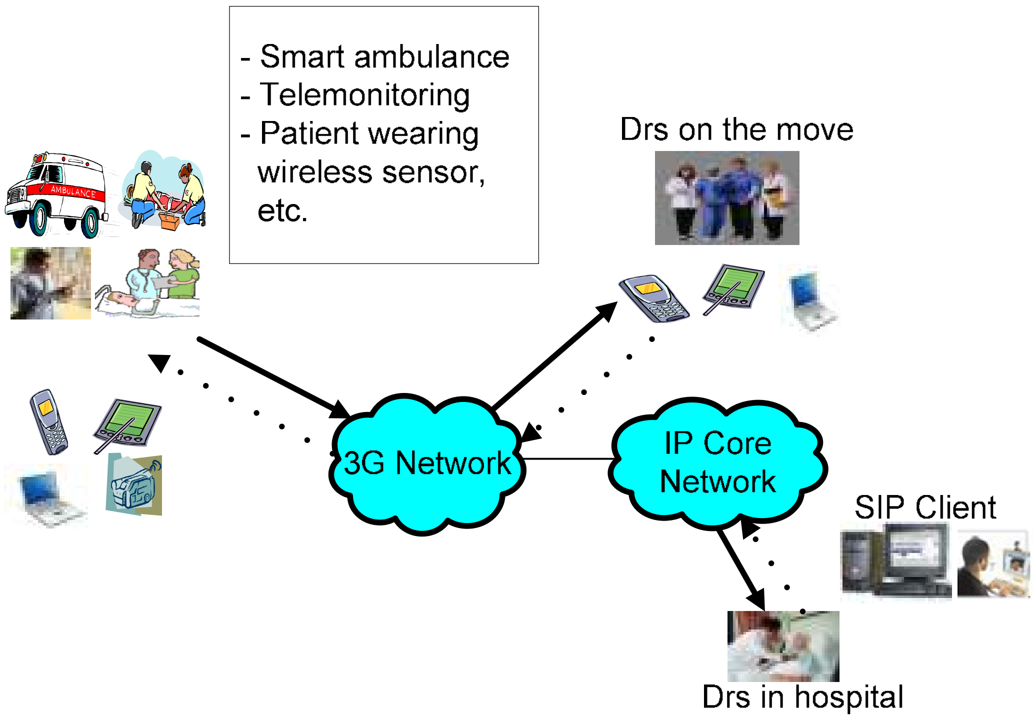


Figure 8.3. Various scenarios utilizing 3G wireless network

Fig. 8.3 outlines a number of scenarios of 3G wireless network applications. Consider a situation where a roadside accident has occurred. Emergency staff in the ambulance would contact via videoconferencing to the hospital from the ambulance, hence delivering healthcare instantly. It may also be that the condition of a high risk patient who wears a wireless sensor has worsened. The wireless sensor that patient is wearing will transmit patient's vital signs to the hospital/ to a doctor on the move. In either case the 3G network is used to establish communication link between the incident site and the health professionals whether on the move or in the hospital via the IP core network. Transmission of data, medical images and live video of the patient can enormously help the junior doctor on site to help improve the condition of the patient. A number of devices e.g. a mobile phone, Personal Digital Assistant (PDA), ipad or a laptop can be used to initiate the mobile voice/video communication.

Resource-efficient QoS requirements in this Scenario

High quality medical images such a single chest radiograph may require from 40 to 50 Mbytes [174]. With a circuit switched 64Kb/s the restricted bandwidth compression techniques would be crucial and a compromise will be needed with the lossy compression techniques currently available. Similarly, video packet/block/frame loss etc. causes video artefacts such as blockiness, blurriness etc. resulting in loss of some data rendering it impossible for the medical professional to successfully interpret the results. An accurate model can measure quality according to the requirement of the healthcare professional.

Security and privacy of medical data is also of concern [175]. Procedures should be put in place to ensure that only authorized personnel have access to the data with encryption etc.

Future trends dictate that mobile handsets would be dual mode – in hotspots utilize the wireless networks IEEE 802.11-WLAN instead of 3G networks. IEEE 802.11- WLAN offers high bandwidth connections at low cost but in limited range compared to 3G network that has an increased area of coverage [176].

The main goals of e-healthcare is to increase the accessibility of healthcare professionals, increase the quality of and continuity of care to patients, focus on preventive medicine through early intervention to name a few and hence reduce the overall cost of healthcare and enhance the quality of care. The users QoS requirement will be the driving force for e-health services to take off. The healthcare application will dictate the accuracy of data e.g. for high emergency situation where human lives are at risk QoS is vital whereas for low-medium situations that are not real-time sensitive a certain amount of compromise in the QoS is acceptable. The requirements for the performance of service (such as delay, jitter), expected security level, acceptable price, etc. will be dictated by the healthcare professionals. The exchange of information over 3G network will allow healthcare professionals to have quick access to patient information such as x-rays and CAT scans, as well as notify doctors (both in the hospital and on the move) immediately in case of an emergency – at home as in the mobihealth project [177, 178]-[179] by sending signals via a body area network wore by the patient or outside the home in a remote location. Workflow in outpatient clinics can be improved via short messaging and multimedia messaging service. Even the healthcare professional's rota system can be improved with the use of 3G wireless network [180]. In summary, the information exchange could be a video call, video streaming, web browsing, video conference, short messaging service, multimedia messaging service, etc. to name a few dictated by the situation.

Also the future 3G networks i.e. 4G and beyond will be completely packet based and hence potentially open the bandwidth restriction from 64kb/s (currently for circuit switched video call) to 20Mb/s and beyond (e.g. Long Term Evolution).

8.5 Conclusions

Motivated by the exponential growth of video applications over wireless access networks the project was initiated to investigate the interactions of various QoS parameters and to develop models that are light weight and can accurately and efficiently predict video quality in a non-intrusive manner thus avoiding time consuming and expensive subjective tests.

Non-linear regression models and neural network models have been developed and further applied to two main applications, namely, optimization of content and network provisioning and in QoS control. A detailed understanding of the relationships between video quality, wireless access networks (both WLAN and UMTS) impairments, impairments associated with the codec and content itself have been provided in order to derive the models properly.

The novelty in this work are the new non-linear regression-based and neural network-based models that combine parameters associated with the encoder, access network and content type and predict video quality non-intrusively, a new content classification method, methods to optimize video contents and network resources and a new QoS fuzzy control scheme .

For most part of of this research an understanding of the problems for video quality prediction for video applications over wireless access networks, objective and subjective video quality measurement, end-to-end video quality prediction, optimization and control. The outcomes of this research can be used as building blocks for future work in this area. However, their performance need to be further confirmed with physical systems, real test beds and large scale networks and for other applications before they are made available for commercial applications and before a realistic implementation can be made for QoS sensitive scenarios, such as the e-healthcare emergency.

References

[1] "ITU T Rec. P.910, "Subjective video quality assessment methods for multimedia applications"," Geneva, Sep. 1999.

[2] "Video Quality Experts Group (VQEG), Multimedia group test plan, "Final report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment", Version 1.2, March, 2008, **http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI."**

[3] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication,* vol. 19, no. 2, pp. 121-132, 2004.

[4] C. Ke, C. Lin, C. Shieh *et al.*, "A novel realistic simulation tool for video transmission over wireless network," *Proc. of the IEEE Int. Conf. on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC'06)*, 2006.

[5] VQM, "NTIA/ITS model **http://www.its.bldrdoc.gov/vqm/** software.php."

[6] "ITU-T Rec. J. 247 (08/08) "Objective perceptual multimedia video quality measurement in the presence of of a full reference."

[7] H. Koumaras, A. Kourtis, C. Lin *et al.*, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences," *Third international conference on Networking and Services*, 19-25 June, 2007.

[8] A. Eden, "No-Reference Image Quality Analysis for Compressed Video Sequences," *Broadcasting, IEEE Transactions on,* vol. 54, no. 3, pp. 691-697, 2008.

[9] Q. Huynh-Thu, and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," *IEEE Transactions on Broadcasting,* vol. Vol. 54, no. No. 3, pp. pp 641-651, 2008.

[10] R. Feghali, F. Speranza, D. Wang *et al.*, "Video quality metric for bitrate control via joint adjustment of quantization and frame rate," *IEEE Transactions on Broadcasting,* vol. 53, no. 1, March, 2007.

[11] M. Ries, O. Nemethova, and M. Rupp, "Video quality estimation for mobile H.264/AVC video streaming," *Journal of Communications,* vol. 3, no. 1, pp. 41-50, 2008.

[12] P. Calyam, E. Ekicio, C. Lee *et al.*, "A gap-model based framework for online VVoIP QoE measurement," *Journal of Communications and Networks,* vol. Vol. 9, no. No.4, pp. pp. 446-56, Dec. 2007.

[13] S. Tao, J. Apostolopoulos, and R. Guerin, "Real-Time Monitoring of Video Quality in IP Networks," *Proc. of ACM NOSSDAV*, 2005.

[14] S. Mohamed, and G. Rubino, "A study of real-time packet video quality using random neural networks," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 12, no. 12, pp. 1071-1083, 2002.

[15] G. Zhai, J. Cai, W. Lin *et al.*, "Cross Dimensional Scalable Video Adaptation via User-End Perceptual Quality Assessment," *Broadcasting, IEEE Transactions on,* vol. 54, no. 3, pp. 719-727, 2008.

[16] K. Seshadrinathan, and A. C. Bovik, "Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos," *Image Processing, IEEE Transactions on,* vol. 19, no. 2, pp. 335-350.

[17] G. W. Cermak, "Subjective video quality as a function of bit rate, frame rate, packet loss, and codec," *2009 International Workshop on Quality of Multimedia Experience, QoMEx 2009.* pp. 41-46.

[18] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan *et al.*, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 20, no. 4, pp. 587-599.

[19] K. Seshadrinathan, R. Soundararajan, A. C. Bovik *et al.*, "Study of Subjective and Objective Quality Assessment of Video," *Image Processing, IEEE Transactions on,* vol. 19, no. 6, pp. 1427-1441.

[20] "ITU-T SG 9, Q 12/9, Hybrid perceptual/bitstream models (**http://www.itu.int/itu-t/workprog/wp_item.aspx?isn=6299)."**

[21] J. Klaue, B. Tathke, and A. Wolisz, "Evalvid – A framework for video transmission and quality evaluation," *Proc. Of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois, USA*, pp. 255-272, Sept. 2003.

[22] A. Khan, Z. Li, L. Sun *et al.*, "Audiovisual Quality Assessment for 3G Networks in Support of E-Healthcare Services, Proceedings of the 3rd International Conference on Computational Intelligence in Medicine and Healthcare(CIMED 2007),July 25-27, Plymouth, UK, 2007."

[23] A. Khan, L. Sun, E. Ifeachor *et al.*, "Impact of RLC losses on quality prediction for H.264 video over UMTS networks, IEEE International Conference on Multimedia and Expo (ICME), 19-23 July, Singapore." pp. 702-707.

[24] A. Khan, L. Sun, and E. Ifeachor, "Content-based video quality prediction for MPEG4 video streaming over wireless networks," *Journal of Multimedia,* vol. 4, no. 4, pp. 228-239, 2009.

[25] A. Khan, "Video Quality Prediction Models based on Video Content Dynamics for H.264 Video over UMTS Networks," *International Journal of Digital Multimedia Broadcasting,* vol. vol. 2010, Article ID 608138, doi:10.1155/2010/608138, pp. 17, 2010.

[26] A. Khan, "Video Quality Assessment as Impacted by Video Content over Wireless Networks " *International Journal on Advances in Networks and Services,* vol. 2, no. 2&3, 2009.

[27] A. Khan, L. Sun, and E. Ifeachor, "Content Classification Based on Objective Video Quality Evaluation for MPEG4 Video Streaming over Wireless Networks " in World Congress on Engineering, ICWN Track, London, UK, 2009.

[28] A. Khan, L. Sun, and E. Ifeachor, "Impact of video content on video quality for video over wireless networks," *Proceedings of the 5th International Conference on Autonomic and Autonomous Systems, ICAS 2009.* pp. 277-282.

[29] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," in Proceedings of the 2009 IEEE international conference on Communications, Dresden, Germany, 14-18 June, 2009.

[30] A. Khan, L. Sun, E. Ifeachor *et al.*, "Video Quality Prediction Model for H.264 Video over UMTS Networks and Their Application in Mobile Video Streaming,

IEEE International Conference on Communications (ICC), Cape Town, South Africa, 23-27 May."

[31] A. Khan, L. Sun, and E. Ifeachor, “Content-based video quality prediction model for H.264 video and its application in QoE-driven sender bitrate adaptation over UMTS networks,” *IEEE Transactions on Multimedia*, In review.

[32] A. Khan, L. Sun, and E. Ifeachor, “QoE Prediction Model and its Application in Video Quality Adaptation over UMTS Networks,” *IEEE Transactions on Multimedia (Accepted for publication)*.

[33] A. Khan, L. Sun, and E. Ifeachor, “An ANFIS-Based Hybrid Video Quality Prediction Model for Video Streaming over Wireless Networks,” *Proceedings of the second International Conference on Next Generation Mobile Applications, Services, and Technologies (NGMAST)*, 16-19 Sept., 2008.

[34] A. Khan, L. Sun, and E. Ifeachor, “Learning models for video quality prediction over wireless local area network and universal mobile telecommunication system networks,” *Communications, IET,* vol. 4, no. 12, pp. 1389-1403.

[35] A. Khan, L. Sun, E. Ifeachor *et al.*, "An ANFIS-based Hybrid Quality Prediction Model for H.264 Video over UMTS Network, IEEE Workshop on Communication, Quality and Reliability (CQR), Vancouver, Canada, 8-10 June."

[36] A. Khan, L. Sun, E. Jammeh *et al.*, “Quality of experience-driven adaptation scheme for video applications over wireless networks,” *Communications, IET,* vol. 4, no. 11, pp. 1337-1347.

[37] A. Khan, L. Sun, E. Jammeh *et al.*, “Content classification-based and QoE-driven video send bitrate adaptation scheme,” in Proceedings of the 5th International ICST Mobile Multimedia Communications Conference, London, United Kingdom, 7-9 Sept., 2009.

[38] E. Jammeh, I. Mkwawa, A. Khan *et al.*, “Quality of experience (QoE) driven adaptation scheme for voice/video over IP,” *Telecommunication Systems*, pp. 1-13.

[39] A. Khan, I. Mkwawa, L. Sun *et al.*, “QoE-driven Sender Bitrate Adaptation Scheme for Video Applications over IP Multimedia Subsystem,” *IEEE ICC 2011, Kyoto, Japan, 5-9 June*, 2011.

[40] “BT-500-11: Methodology for subjective assessment of the quality of Television Picture. International Telecommunication Union.”

[41] “Subjective MOS @ **www.tech.plym.ac.uk/spmc/staff/akhan/mos_scores.html.”**

[42] A. Khan, L. Sun, J. Fajardo *et al.*, “Impact of End Devices on Subjective Video Quality Assessment for QCIF Video Sequences,” *QoMEX, 7-9 Sept., Mechelen, Belgium*, 2011.

[43] “NS2, **http://www.isi.edu/nsnam/ns/.”**

[44] “OPNET for Research **http://www.opnet.com.”**

[45] “Ffmpeg, **http://sourceforge.net/projects/ffmpeg.”**

[46] “JM H.264 Software, **http://iphome.hhi.de/suehring/tml/.”**

[47] “IEEE 802.11n standards **http://standards.ieee.org/getieee802/802.11.html.”**

[48] “3GPP TS 25.322, “Third Generation Partnership Project: Technical Specification Group Access network; Radio Link Control (RLC Specification (Release 5).”

[49] “RTP: A Transport Protocol for Real Time Applications RFC3550.”

[50] T. Freidman, R.Caceres, and A. Clark, "RTP Control Protocol Extended Reports (RTCP XR)," Nov. 2003.

[51] J. E. Ebert, and A. Willig, "A Gilbert-Elliot bit error model and the efficient use in packet level simulation " *Technical Report, TKN-99-002, Technical University of Berlin*, March, 1999.

[52] H. S. Wang, and N. Moayeri, "Finite state Markov channel - a useful model for radio communication channel," *IEEE Transactions on Vehicular Technology,* vol. 44, no. 1, pp. 163-171, Feb. 1995.

[53] "Enhanced UMTS Radio Access Network Extensions for ns-2 (E.U.R.A.N.E) **http://eurane.ti-wmc.nl/eurane/."**

[54] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Systems Technical Journal,* vol. 39, pp. 1253-1265, Sept. 1960.

[55] M. Ghanbari, "Standard codecs: image compression to advanced video coding," *IEE Telecommunication Series 49*, 2003.

[56] B. Girod, "EE368b Image and Video Compression, **http://www.stanford.edu/class/ee368b/Handouts/19-CodingStandards.pdf."**

[57] J. Mitchell, and W. Pennebaker, "MPEG Video: Compression Standard, Chapman and Hall, ISBN 0412087715," 1996.

[58] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions," *SPIE Conference on Applications of Digital Image Processing XXVII*, Aug. 2004.

[59] "Subjective VoIP tests @ **http://www.tech.plym.ac.uk/spmc/people/lfsun/mos/Default.htm."**

[60] C. Ke, C. Lin, and C. Shieh, "Evaluation of streaming MPEG video over wireless channels," *Journal of mobile multimedia,* vol. Vol. 3, no. No. 1, pp. pp. 047-064, 2007.

[61] T. Pliakas, G. Kormentzas, and C. Skianis, "Scalable video streaming traffic delivery in IP/UMTS networking environments," *Journal of Multimedia,* vol. 2, no. 2, 2007.

[62] I. Politis, T. Dagiuklas, and S. Kotsopoulos, "Study of the QoS of video traffic over integrated 3G-WLAN systems," *Mobimedia'06, Alghero, Sardinia, Italy, 18-20 September*, 2006.

[63] F. Angelis, I. Habib, F. Davide *et al.*, "Intelligent content aware services in 3G wireless networks," *IEEE Journal on selected areas in Communications,* vol. 23, no. 2, Feb., 2005.

[64] A. Salkintzis, D. Skyrianoglou, and N. Passas, "Seamless multimedia QoS across UMTS and WLANs," *Vehicular Technology Conference, 30th May-1st June* 2005.

[65] A. Lo, G. Heijenk, and I. Niemegeers, "Performance evaluation of MPEG4 video streaming over UMTS networks using an integrated tool environme," *International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS, Philadelphia, PA, USA, 24 - 28 July* pp. 676-682, 2005.

[66] A. Alexiou, C. Bouras, and V. Igglesis, "Multimedia transmission over third generation cellular network," *Proc. of the 10th Int. IFIP Int. conference, Colmar, France, 25-27* 2005.

[67] V. Vassiliou, P. Antoniou, I. Giannakou *et al.*, “Requirements for the transmission of streaming video in mobile wireless networks,” *Part I-II: 16th international conference, ICANN, Athens, Greece, 10-14 September*, 2006.

[68] A. Alexiou, D. Antonellis, and C. Bouras, “Adaptive and reliable video transmission over UMTS for enhanced performance
” *Int. Journal of Communication Systems* vol. 20, no. 1, pp. 65-81, Jan., 2007.

[69] O. Nemethova, W. Karner, A. Al-Moghrabi *et al.*, “Cross layer detection for H264 video over UMTS,” *Proc. of Int. Wireless Summit, Aolborg, Denmark, 18-22 Dec.* , 2005.

[70] S. Dogan, S. Eminsoy, A. Sadka *et al.*, “Video content adaptation using transcoding for enabling UMA over UMTS,” *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'2004), Lisbon, Portugal, April* 2004.

[71] S. Song, Y. Won, and I. Song, “Empirical study of user perception behavious for mobile streaming,” *Proc. of ACM, Juan-les-Pins, France, 1-6 Dec.* , 2002.

[72] H.Knoche, J. McCarthy, and A. Sasse, “Can small be beautiful? Assessing image resolution requirements for mobile TV,” *Proc. of ACM, Singapore, 6-11 Nov.*, 2005.

[73] S. Jumisko-Pyykko, and J.Hakkinen, “Evaluation of subjective video quality of mobile devices,” *Proc. of ACM, Singapore, 6-11 Nov.*, 2005.

[74] A. Eichhorn, and P. Ni, "Pick Your Layers Wisely - A Quality Assessment of H.264 Scalable Video Coding for Mobile Devices, Proc. of ICC '09, 14-18 June, Dresden, Germany, 2009."

[75] N. Cranley, and M. Davis, “Performance evaluation of video streaming with background traffic over IEEE 802.11 WLAN neyworks,” *Proc. of ACM, WMuNeP, 13 Oct., Montreal, Canada*, 2005.

[76] Y. Koucheryavy, D. Moltchanov, and J. Harju, “Performance Evaluation of Live Video Streaming Service in 802.11b WLAN Environment under Different Load Conditions,” *Interactive Multimedia on Next Generation Networks, First International Workshop on Multimedia Interactive Protocols and Systems, MIPS 2003, Napoli, Italy, 18-21 Nov.*, 2003.

[77] Z. Guangtao, L. Weisi, C. Jianfei *et al.*, "Cross-dimensional perceptual quality assessment for low bitrate video, IEEE Transactions on Multimedia, vol.10, issue 7, Nov., 2008." pp. 1316-1324.

[78] T. D. D. Koning, P. Velthoven, H. Knoche *et al.*, “Of MOS and men: bridging the gap between objective and subjective qualitymeasurements inMobile TV,” *Multimedia on Mobile Devices, Proceedings of the SPIE, p. 65070P*, 2007.

[79] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan *et al.*, “LIVE Wireless Video Quality Assessment Database", URL:**http://live.ece.utexas.edu/research/quality/live_wireless_video.html**. ,” 2009.

[80] “Raw Videos **http://trace.eas.asu.edu/yuv/index.html.”**

[81] “MP4BOX, "**http://www.videohelp.com/tools/mp4box".”**

[82] W. Karner, O. Nemethova, P. Svoboda *et al.*, “Link Error Analysis and Modeling for Video Streaming Cross-Layer Design in Mobile Communication Networks,” *ETRI Journal,* vol. 29, no. 5, Oct. 2007.

[83] W. J. Krzanowski, “Principles of multivariate analysis,” *Clarendon press, Oxford*, 1988.

[84] Y. Hochberg, and A. C. Tamhane, “Multiple Comparison Procedures,” Wiley, 1987.

[85] M. Ries, C. Crespi, O. Nemethova *et al.*, "Content based video quality estimation for H.264/AVC video streaming," *IEEE Wireless Communications and Networking Conference, WCNC.* pp. 2670-2675.

[86] L. Yu-xin, K. Ragip, and B. Udit, "Video classification for video quality prediction," *Journal of Zhejiang University Science A,* vol. 7, no. 5, 2006.

[87] H. Koumaras, and A. Kourtis, "Video Quality Prediction based on the Spatial and Temporal Classification of the Uncompressed Content, IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, 3-7 Sept., 2007 ". pp. 1-5.

[88] H. Koumaras, and A. Kourtis, "Video quality prediction based on the spatial and temporal classification of the uncompressed content," *The 18th IEEE Int. Symoposium on Personal, Indoor and Mobile Radio Communication (PIMRC)*, 2007.

[89] Y. Kato, A. Honda, and K. Hakozaki, "An analysis of relationship between video contents and subjective," *Proc. Of the 19th Int. Conf. On Advanced Information Networking and Applications (AINA)*, 2005.

[90] Y. Suda, K. Yamori, and Y. Tanaka, "Content clustering based on users' subjective evaluation," *6th Asia-Pacific symposium on Information and Telecommunication Technologies ASPITT,* vol. 9, no. 10, pp. 177–182, 2005.

[91] S. d. Toit, A. Steyn, and R. Stumpf, "Cluster analysis, Handbook of graphical exploratory data analysis," *ed. S.H.C. du Toit, Springer-Verlag, New York*, pp. 73-104, 1986.

[92] J. Wei, "Video content classification based on 3-d Eigen analysis," *IEEE transactions on image processing,* vol. 14, no. 5, May 2005.

[93] G. Gao, J. J, J. Liang *et al.*, "PCA-based approach for video scene change detection on compressed video," *Electronics letters, 23rd Nov. ,* vol. 42, no. 24, 2006.

[94] N. Cranley, L. Murphy, and P. Perry, "Content-Based Adaptation of Streamed Multimedia " *Management of Multimedia Networks and Services, Lecture Notes in Computer Science* vol. 3271, pp. 151-197, 2004.

[95] H. Koumaras, J. Arnaud, D. Négru *et al.*, "An Experimental Approach of Video Quality Level Dependence on Video Content Dynamics," *In 5th International Mobile Multimedia Communications Conference, Mobimedia* 2009.

[96] P. Chen, "Fully scalable subband/wavelet coding, Ph.D. dissertation, Rensselaer Polytechnic Inst., Troy, NY, May," 2003.

[97] R. Ohm, M. Schaar, and J. Woods, "Interframe wavelet coding— Motion picture representation for universal scalability," *Signal Process.Image Commun., vol. 19, no. 9, pp. 877–908, Oct.* , 2004.

[98] M. Al-Mualla, C. Canagarajah, and D. Bull, "Video coding for mobile communications " *Academy Press – An imprint of Elsevier Science*, 2002.

[99] S. Minami, and A. Zakhor, "Optimization approach for removing blocking effects in transform coding," *IEEE Trans. On Circuits and Systems for Video Technology, Vol. 5, No. 7, pp. 74-82*, 1995.

[100] R. J. Hyndman, "The problem with Sturges' rule for constructing histograms," *Monash University* 1995.

[101] D. S. Hands, "Video QoS enhancement using perceptual quality metrics," *BT Technology Journal, April* vol. 23, no. 2, pp. 208-216, 2005.

[102] Y. Fei, and C. En, "Reduced-reference metric design for video quality measurement in wireless application, 11th IEEE International Conference on Communication Technology (ICCT), 10-12 Nov. 2008 ". pp. 641-644.

[103] T. Oelbaum, and K. Diepold, "Building a Reduced Reference Video Quality Metric with Very Low Overhead Using Multivariate Data Analysis," *The Journal on Systemics, Cybernetics and Informatics (JSCI),* vol. 6, no. 5, pp. 81-86, 2008.

[104] M. Naccari, M. Tagliasacchi, and S. Tubaro, "No-reference video quality monitoring for H.264/AVC coded video," *IEEE Transactions on Multimedia, August,* vol. 11, no. 5, 2009.

[105] J. Li, G. Lu, G. Chen *et al.*, "A method of video quality assessment based on the sensitive reqion," *World Congress on Computer Science and Information Engineering*, 2008.

[106] S. P´echard, D. Barba, and P. Le Callet, "Video quality model based on a spatio-temporal features extractions for H.264 coded HDTV sequences," *Proc. Picture Coding Symposium, Nov.*, 2007.

[107] Y. Kawayoke, and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," *Proceedings - International Conference on Image Processing, ICIP.* pp. 385-388.

[108] S. Winkler, and F. Dufaux, "Video quality evaluation for mobile streaming applications," *SPIE Conf Visual Communication and Image Processing, Lugano, Switzerland*, 2003.

[109] P. Marziliano, F. Dufaux, S. Winkler *et al.*, "A no-reference perceptual blur metric, International Conference on Image Processing, 2002."

[110] A. R. Reibman, V. A. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia,* vol. 6, no. 2, pp. 327-334, April, 2004.

[111] K.-C. Yang, C. C. Guest, K. El-Maleh *et al.*, "Perceptual temporal quality metric for compressed video," *IEEE Transactions on Multimedia,* vol. 9, no. 7, pp. 1528-1535, Nov., 2007.

[112] M. Ries, O. Nemethova, B. Badic *et al.*, "Assessment of H.264 coded panorama sequences, 1st International Conference on Multimedia Services Access Networks, MSAN '05. 2005." pp. 6-9.

[113] Z. Wang, S. Banerjee, and S. Jamin, "Studying streaming video quality: From an application point of view," *Proc. of ACM, Berkeley, California, USA, 2-8 Nov.* , 2003.

[114] A. G. Davis, D. Bayart, and D. S. Hands, "Hybrid no-reference video quality prediction," *2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB 2009*.

[115] M. Soleymani, G. Chanel, J. Kierkels *et al.*, "Affective Characterization of Movie Scenes Based on Multimedia Content Analysis and User's Physiological Emotional Responses." pp. 228-235.

[116] K. Yamagishi, and T. Hayashi, "Opinion model for estimating video quality of videophone services," *Proc. of IEEE Globecom, San Francisco, USA, 27th Nov to 1st Dec.*, 2006.

[117] S. Winkler, and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *Broadcasting, IEEE Transactions on,* vol. 54, no. 3, pp. 660-668, 2008.

[118] J. Hu, S. Choudhury, and J. D. Gibson, "PSNRr,f-MOSr: An Easy-To-Compute Multiuser Perceptual Video Quality Measure," *First International Workshop on Quality of Multimedia Experience (QoMEX), 29-31 July, San Diego, California, USA*, 2009.

[119] "ITU-T Rec. J.148, Requirements for an objective perceptual multimedia quality model"," May 2003.

[120] S. Jumisko-Pyykkö, and M. M. Hannuksela, "Does Context Matter in Quality Evaluation of Mobile Television," *Proceedings of 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2008), Amsterdam, Netherland, 2-5 Sept.*, 2008.

[121] "Degraded videos over WLAN "**www.tech.plymouth.ac.uk/spmc/staff/akhan/video_sequences/videoclips.html".**"

[122] "NTIA/ITS model **http://www.its.bldrdoc.gov/vqm/software.php."**

[123] G. Rubino, and M. V. Irisa, "A new approach for the prediction of end-to-end performance of multimedia streams," *Proc. of the First Int. Conf. on the Quantitative Evaluation of Systems (QEST), Enschede, The Netherlands, 27-30 sept.*, 2004.

[124] G. Rubino, M. Varela, and J.-M. Bonnin, "Controlling multimedia QoS in the future home network using the PSQA metric," *The Computer Journal, March,* vol. 49, no. 2, 2006.

[125] M. Ries, J. Kubanek, and M. Rupp, "video quality estimation for mobile streaming applications with neuronal networks," *Proc. of MESAQIN conference, Prague, Czech Republic, 5-6 June*, 2006.

[126] C. Jihwan, L. Kwon, and L. Chulhee, "No-reference video quality measurement using neural networks, 16th International Conference on Digital Signal Processing, 5-7 July, 2009 ". pp. 1-4.

[127] J. Xiuhua, M. Fang, X. Jiangbo *et al.*, "No-Reference Perceptual Video Quality Measurement for High Definition Videos Based on an Artificial Neural Network." pp. 424-427.

[128] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment," *Neural Networks, IEEE Transactions on,* vol. 17, no. 5, pp. 1316-1327, 2006.

[129] A. Abdennour, "Evaluation of neural network architectures for MPEG-4 video traffic prediction," *Broadcasting, IEEE Transactions on,* vol. 52, no. 2, pp. 184-192, 2006.

[130] R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man, and Cybernetics, May,* vol. 23, no. 3, pp. 665-685, 1993.

[131] "MATLAB **http://www.mathworks.co.uk/."**

[132] W. Yong, M. van der Schaar, C. Shih-Fu *et al.*, "Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 15, no. 10, pp. 1270-1279, 2005.

[133] O. D. Onur, and A. Aydm Alatan, "Video Adaptation Based on Content Characteristics and Hardware Capabilities, Second International Workshop on Semantic Media Adaptation and Personalization, 17-18 Dec., 2007."

[134] H. Ming-Ho, C. Yi-Wen, C. Hua-Tsung *et al.*, "Content-aware video adaptation under low-bitrate constraint," *EURASIP J. Adv. Signal Process,* vol. 2007, no. 2, pp. 27-27, 2007.

[135] G. M. Marcelo, and G. Rudinei, "Live video adaptation: a context-aware approach," in Proceedings of the 11th Brazilian Symposium on Multimedia and the web, Pocos de Caldas - Minas Gerais, Brazil, 2005.

[136] A. Carreras, J. Delgado, E. Rodriguez *et al.*, “A platform for context-aware and digital rights management enabled content adaptation,” *IEEE Multimedia*, no. 2, pp. 74-89, April-June, 2010.

[137] F. De Angelis, I. Habib, F. A. M. Davide *et al.*, “Intelligent content aware services in 3G wireless networks,” *Selected Areas in Communications, IEEE Journal on,* vol. 23, no. 2, pp. 221-234, 2005.

[138] P. Papadimitriou, and V. Tsaoussidis, "A Rate Control Scheme for Adaptive Video Streaming Over the Internet, IEEE International Conference on Communications, 24-28 June, Glasgow, UK, 2007 ". pp. 616-621.

[139] Z. Qian, Z. Wenwu, and Z. Ya-Qin, “Resource allocation for multimedia streaming over the Internet,” *Multimedia, IEEE Transactions on,* vol. 3, no. 3, pp. 339-355, 2001.

[140] A. G. Alexiou, C. J. Bouras, and V. G. Igglesis, "A decision feedback scheme for multimedia transmission over 3G mobile networks, Second Int. Conf. on Wireless and Optical Communications Networks, WOCN, 6-8 March, 2005."

[141] G. Harinath, C. Hyukjune, S. Naveen *et al.*, "Rate adaptation for video telephony in 3G networks, Packet Video, 12-13 Nov., 2007."

[142] A. Antonios, A. Dimitrios, and B. Christos, “Adaptive and reliable video transmission over UMTS for enhanced performance: Research Articles,” *Int. J. Commun. Syst.,* vol. 20, no. 1, pp. 65-81, 2007.

[143] D. Kim, and K. Jun, “Dynamic bandwidth allocation scheme for video streaming in wireless cellular networks,” *IEICE Trans. Commun., Feb. ,* vol. E89-B, no. 2, 2006.

[144] S. Gnavi, M. Grangetto, E. Magli *et al.*, “Comparison of rate allocation strategies for H.264 video transmission over wireless lossy correlated networks,” in Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 1, 2003.

[145] T. Stockhammer, D. Kontopodis, and T. Wieg, “Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment ” *International Packet Video Workshop*, 2002.

[146] G. Nur, H. Arachchi, S. Dogan *et al.*, “Evaluation of quality scalability layer selection for bit rate adaptation of scalable video content,” *10th Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS '09.* , 6-8 May 2009

[147] T. Pliakas, G. Kormentzas, and C. Skianis, “Scalable video streaming traffic delivery in IP/UMTS networking environments,” *Journal of Multimedia, April,* vol. 2, no. 2, pp. 37-46, 2007.

[148] P. Antoniou, V. Vassiliou, and A. Pitssillides, “Delivering adaptive scalable video over the wireless networks,” *1st ERCIM Workshop on eMobility, Coimbra, Portugal, May 21*, pp. 23-24, 2007.

[149] P. Antoniou, V. Vassiliou, and A. Pitsillides, “ADIVIS: A novel adaptive algorithm for video streaming over the Internet,” *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’07)*, 2007.

[150] T. Hayashi, G. Kawaguti, J. Okamota *et al.*, “Subjective quality estimation model for video streaming services with dynamic bit-rate control,” *IEICE Trans. Commun., Feb.,* vol. E89-B, no. 2, pp. 297-303, 2006.

[151] E. Jammeh, M. Fleury, and M. Ghanbari, “Fuzzy logic congestion control of transcoded video streaming without packet loss feedback,” *IEEE Transactions*

on Circuits and Systems for Video Technology, vol. 18, no. 3, pp. 387-393, 2008.
[152] D. Mukherjee, E. Delfosse, K. Jae-Gon *et al.*, “Optimal adaptation decision-taking for terminal and network quality-of-service,” *Multimedia, IEEE Transactions on,* vol. 7, no. 3, pp. 454-462, 2005.
[153] C. Timmerer, V. H. Ortega, J. M. Gonzalez *et al.*, "Measuring quality of experience for MPEG-21-based cross-layer multimedia content adaptation." pp. 969-974.
[154] C. Eduardo, V. Luis, Mar *et al.*, “QoS mapping and adaptation control for multi-user sessions over heterogeneous wireless networks,” in Proceedings of the 3rd international conference on Mobile multimedia communications, Nafpaktos, Greece, 2007.
[155] B. Lewcion, S. Moller, M. Waltermann *et al.*, “Methods for multimedia service adaptation in next generation networks,” *Proc. of the ETSI Workshop on Effects of Transmission Peformance on Multimedia QoS, June* 2008.
[156] F. Agboma, and A. Liotta, “QoE-aware QoS management, 6th Int. Conf. on Advances in Mobile computing and Multimedia, 24-26 Nov. 2008.”
[157] G. M. Muntean, P. Perry, and L. Murphy, “A new adaptive multimedia streaming system for all-IP multi-service networks,” *Broadcasting, IEEE Transactions on,* vol. 50, no. 1, pp. 1-10, 2004.
[158] B. Ciubotaru, G. Muntean, and G. Ghinea, “Objective assessment of region of interest-aware adaptive multimedia streaming quality,” *IEEE Transactions on Broadcasting, June,* vol. 55, no. 2, 2009.
[159] F. Pereira, “Sensations, perceptions and emotions towards quality of experience evaluation for consumer electronics video adaptations,” *Proc. of Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, Jan*, 2005.
[160] M. Rezaei, M. M. Hannuksela, and M. Gabbouj, “Semi-Fuzzy Rate Controller for Variable Bit Rate Video,” *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 18, no. 5, pp. 633-645, 2008.
[161] “Telchemy application notes, “Understanding of video quality metrics”, Telchemy, Feb. 2008. **http://www.telchemy.co.uk.”**
[162] H. Takagi, "Application of neural networks and fuzzy logic to consumer products, International Conference on Industrial Electronics, Control, Instrumentation, and Automation, 1629-1633, vol.3,9-13 Nov.,1992."
[163] T. Freidman, R.Caceres, and A. Clark, “RTP Control Protocol Extended Reports (RTCP XR), November,” 2003.
[164] M. Handley, S. Floyd, J. Widmer *et al.*, “RFC3448: TCP-Friendly Rate Control (TFRC): Protocol Specification. **http://ww.ietf.org/rfc/rfc3448.txt.,”** 2003.
[165] A. Lie, and J. Klaue, “Evalvid-RA: Trace driven simulation of rate adaptive MPEG4 VBR video ” *Multimedia Systems,* vol. 14, no. 1, pp. 33-50, 2008.
[166] “Fokus, “Open ims playground,” Website, 2008, **http://www.fokus.fraunhofer.de/en/fokus_testbeds/open** ims playground/ index.html.”
[167] “Android Platform, “Android-an open handset alliance project,” Website, 2008, **http://code.google.com/android/index.html..”**
[168] I. Mkwawa, E. Jammeh, L. Sun *et al.*, “Feedback-free early VoIP quality adaptation scheme in next generation network,” *IEEE Globecom, 6-10 December, Miami, Florida, USA*, 2010.

[169] S. Winkler, and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE transactions on multimedia,* vol. 8, no. 5, pp. 973-980, 2006.

[170] N. Kitawaki, "Multimedia quality prediction methodologies for advanced mobile and IP-based telephony," *IEICE Trans. Commun., Feb.,* vol. 89-B, no. 2, 2006.

[171] S. Tasaka, H. Yoshimi, and A. Hirashima, "The effectiveness of a QoE-based video output scheme for audio-video IP transmission," *Proc. of ACM, Vancouver, British Columbia, Canada, 26-31 October* 2008.

[172] W. Chen, J. Fournier, M. Barkowsky *et al.*, "New requirements of subjective video quality assessment methodologies for 3DTV," *Video Processing and Quality Metrics 2010 (VPQM), Scottsdale, United States* 2010.

[173] "**http://www.wirelesshealthcare.co.uk** ".

[174] C. S. Pattichis, E. Kyriacou, S. Voskarides *et al.*, " Wireless telemedicine systems: an overview " *IEEE Antennas & Propogation Magazine,* vol. 44, no. 2, pp. 143-153, 2002.

[175] Z. Li, L. Sun, and E. Ifeachor, "Challenges of mobile ad-hoc grids and their applications in e-healthcare," *Proceedings of Second International Conference on Computational Intelligence in Medicine and Healthcare (CIMED' 2005), July, Lisbon, Portugal.*, 2005.

[176] Y. E. Tan, N. Philip, and R. S. H. Istepanian, "Fragility issues of medical video streaming over 802.11e-WLAN m-health environment," *Proceedings of the 28th IEEE EMBS Annual International Conference, 30th Aug. to 3rd Sept., New York City, USA,* , 2006.

[177] K. Wac, R. Bults, D. Konstantos *et al.*, "Mobile healthcare over 3G networks: the mobihealth pilot system and service," *Proc. Of 2004 Global Mobile Congress, Shangai, 11-13 Oct.* , pp. 71-76, 2004.

[178] V. Jones, A. V. Halteren, R. Bults *et al.*, "MobiHealth: Mobile Healthcare, **http://www.mobihealth.org**, IST – 2001 – 36006.."

[179] E. Jovanov, "Wireless technology and system integration in body area networks for m-health applications," *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, China, September* 2005.

[180] A. Aneiba, S. Hanlon, and H. Shah, "An architecture for mobile web business systems: applied to the healthcare sector," *Proceedings of the Third International Conference on Mobile Business, M-Business* 2004.